

STIC-Biotech/ChemLib

180560

mg

From: Swope, Sheridan  
Sent: Saturday, February 25, 2006 12:04 AM  
To: STIC-Biotech/ChemLib  
Subject: 09/917,376

For 09/917,376, pls intereference search:

SID 1 against the NT and AA databases

SID 3 against the NT and AA databases

SID 4 against the NT and AA databases

SID 5 against the NT and AA databases

RECEIVED  
FEB 27 2006  
STIC

Thanks!

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1656  
Recombinant Enzymes  
571-272-0943 (voice)  
E02B71 Remsen Bld (Office)  
E03C70 Remsen Bld (Mailbox)

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 180560**

**TO: Sheridan Swope**  
**Location: rem/2B71/3C70**  
**Art Unit: 1656**  
**Friday, March 03, 2006**  
**Case Serial Number: 09/917376**

**From: Kristine Hensle**  
**Location: Biotech-Chem Library**  
**REM-1B69**  
**Phone: (571) 272-4161**

**Kristine. Hensle@uspto.gov**

### **Search Notes**

Examiner Swope,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle  
Librarian  
STIC Biotech/Chem Library  
(571)272-4161

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 13:55:49 ; Search time 55.1526 Seconds  
(without alignments)  
1434.576 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5134

Sequence: 1 MDRSENRLTWRRLVLL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5-COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6-COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H-COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE-COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Match | Length | DB | ID                   | Description        |
|------------|-------|-------|--------|----|----------------------|--------------------|
| 1          | 334   | 6.5   | 1751   | 2  | US-09-136-574A-44    | Sequence 44, Appl  |
| 2          | 301.5 | 5.9   | 1749   | 2  | US-09-640-419C-28    | Sequence 28, Appl  |
| 3          | 288   | 5.6   | 96     | 2  | US-09-119-507B-112   | Sequence 112, Appl |
| 4          | 288   | 5.6   | 96     | 2  | US-09-547-693-236    | Sequence 236, Appl |
| 5          | 276   | 5.4   | 1426   | 2  | US-09-136-574A-43    | Sequence 43, Appl  |
| 6          | 260   | 5.1   | 2736   | 2  | US-09-252-901A-30227 | Sequence 30227, A  |
| 7          | 234   | 4.6   | 2137   | 2  | US-09-134-001C-4463  | Sequence 4463, Ap  |
| 8          | 225.5 | 4.4   | 551    | 1  | US-09-033-537A-1     | Sequence 1, Appl   |
| 9          | 221.5 | 4.3   | 258    | 2  | US-10-104-047-3034   | Sequence 3034, Ap  |
| 10         | 221   | 4.3   | 2283   | 2  | US-10-172-502-4      | Sequence 4, Appl   |
| 11         | 219   | 4.3   | 1236   | 2  | US-09-769-787-109    | Sequence 109, Appl |
| 12         | 216.5 | 4.2   | 8991   | 2  | US-08-714-741-32     | Sequence 32, Appl  |
| 13         | 215.5 | 4.2   | 76     | 2  | US-09-547-693-233    | Sequence 233, Appl |
| 14         | 213   | 4.1   | 3892   | 2  | US-09-328-352-5503   | Sequence 5503, Ap  |
| 15         | 210.5 | 4.1   | 521    | 1  | US-08-276-213-3      | Sequence 3, Appl   |
| 16         | 209   | 4.1   | 2315   | 2  | US-09-949-002-314    | Sequence 314, Appl |
| 17         | 209   | 4.1   | 2415   | 2  | US-09-949-002-398    | Sequence 398, Appl |
| 18         | 208.5 | 4.1   | 206    | 2  | US-08-529-055-54     | Sequence 54, Appl  |
| 19         | 208.5 | 4.1   | 493    | 2  | US-09-198-956-10     | Sequence 10, Appl  |
| 20         | 208.5 | 4.1   | 493    | 2  | US-09-198-955A-12    | Sequence 12, Appl  |
| 21         | 208.5 | 4.1   | 493    | 2  | US-09-694-531-12     | Sequence 12, Appl  |
| 22         | 208.5 | 4.1   | 493    | 2  | US-09-670-141-10     | Sequence 10, Appl  |
| 23         | 208.5 | 4.1   | 493    | 2  | US-10-072-152-12     | Sequence 12, Appl  |
| 24         | 205   | 4.0   | 918    | 2  | US-09-200-650E-1     | Sequence 1, Appl   |
| 25         | 204.5 | 4.0   | 1970   | 2  | US-09-538-092-1005   | Sequence 1005, Ap  |
| 26         | 202.5 | 3.9   | 476    | 2  | US-09-339-159B-4     | Sequence 4, Appl   |
| 27         | 200.5 | 3.9   | 700    | 1  | US-07-862-588B-2     | Sequence 2, Appl   |

|    |       |     |      |   |                      |                   |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 200   | 3.9 | 183  | 2 | US-08-529-055-50     | Sequence 50, Appl |
| 29 | 199.5 | 3.9 | 490  | 2 | US-09-109-841-2      | Sequence 2, Appl  |
| 30 | 199.5 | 3.9 | 616  | 2 | US-09-136-574A-47    | Sequence 47, Appl |
| 31 | 199.5 | 3.9 | 1367 | 2 | US-09-487-558B-108   | Sequence 108, App |
| 32 | 197.5 | 3.8 | 1209 | 2 | US-09-949-003-493    | Sequence 493, App |
| 33 | 197.5 | 3.8 | 1209 | 2 | US-09-949-003-494    | Sequence 494, App |
| 34 | 197   | 3.8 | 1060 | 2 | US-08-911-393-2      | Sequence 2, Appl  |
| 35 | 197   | 3.8 | 1060 | 2 | US-09-955-909-2      | Sequence 2, Appl  |
| 36 | 192.5 | 3.7 | 271  | 2 | US-09-248-796A-21050 | Sequence 21050, A |
| 37 | 191.5 | 3.7 | 412  | 1 | US-08-313-288B-18    | Sequence 18, Appl |
| 38 | 191   | 3.7 | 105  | 2 | US-09-547-693-230    | Sequence 230, App |
| 39 | 191   | 3.7 | 2169 | 2 | US-09-949-016-6930   | Sequence 6930, Ap |
| 40 | 190   | 3.7 | 1092 | 2 | US-09-147-408B-15    | Sequence 15, Appl |
| 41 | 189.5 | 3.7 | 1187 | 2 | US-09-949-016-6513   | Sequence 6513, Ap |
| 42 | 189   | 3.7 | 933  | 2 | US-08-293-728-2      | Sequence 2, Appl  |
| 43 | 189   | 3.7 | 933  | 2 | US-09-421-868-2      | Sequence 2, Appl  |
| 44 | 188.5 | 3.7 | 193  | 2 | US-08-529-055-49     | Sequence 49, Appl |
| 45 | 187   | 3.6 | 936  | 2 | US-08-956-171E-5249  | Sequence 5249, Ap |

#### ALIGNMENTS

#### RESULT 1

US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 1751 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 44:

Fabrics Using Truncated

US-09-136-574A-44

Query Match 6.5%; Score 334; DB 2; Length 1751;  
Best Local Similarity 21.2%; Pred. No. 2.5e-13;  
Matches 219; Conservative 125; Mismatches 314; Indels 376; Gaps 56;

Qy 18 SLIAATASFAVAALGVLPITAITASPAHAA-----TTQP---YTWNSVAI 59  
Db 186 SVVAET-----AASLAVASVVIKERNOKAASYLOHAKDLFEFADTTRSDAGTYAATGFY 240  
Qy 60 GCGGFVDGIVFNEGACGILYVRTD-----ICGMVW-----DAANGRWIP 99  
Db 241 TSGGFIDDLGW---AAWLYIATNDSSYLTKABELMSEYANGTWTQCDWDDVRYGFLIM 297  
Qy 100 LLDVWGNNGVNGVSVISAADPINTNKVAAVGM-YTNSWDPNDDGAILRSSDQO----- 152  
Db 298 LAKITGKEL--YKGAVERNLHDWTRITYTPKGMAYLTGM---GSLRYATTAAFLACVY 351  
Qy 153 ATWQ-----ITPLPKLGNMGR-----GMRERLAVDNNNNILYFGAPSG 194  
Db 352 ADWGGCDSNKTKYLINFAKSQIDYALGST--GRSFVVGFGTNPQPHPHRN----- 400  
Qy 195 KGLWRSTDGATWSQMTNEPD-----VGTYIANPTDTTGVQSDIQGVW--VAFDKSS 246  
Db 401 -----AHSSWANSKIPYERHILYLGALVGGPGSDSYNDIDITYVONEVACDYNAG 452  
Qy 247 SLGQASKTIFVGVADPNNPVFWSRDRGATWQAVPGAPTGFPHKGVDPDPVNHVLYIAT-- 304  
Db 453 IVGALAK-----MYQLYGGEPID--DFKAIETPTNDEIFVESKF 489  
Qy 305 SNTGGPYDSSGVDWKFVSVTSGTWTTRISVPVSTDTAN-DYF-----GYSGLTIDRQ 354  
Db 490 GNSQGP--NYTEVISYVNRITGM---PPRVTDKLSFKYFIDITELIQAGYS----- 535  
Qy 355 HPTIMVATOISWMPDTLIFRSTDGATWTRIDWTSYPNRSRYVL-DISASEPWLTFGV 413  
Db 536 -PDVVKVDT-----YYIEGKISGYPVWD--KRNIIYVLVDVSGTK-----I 575  
Qy 414 QNPVPPVSPKLGW---MDERAMIDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPM 469  
Db 576 YPGGEVEHKQAFKISVPGYFWDPTN-DPSYKGLTSQLEKKNYIAAYDNN--L 628  
Qy 470 VKGLETA VNDLSP-----PSGAPLISALDGLGTFHADYVAVPSTFT---SPVFT 519  
Db 629 VMGLEPGAATSTAPTSTPTPTPTPTVTA-----TPTPTPTPTGSP--G 674  
Qy 520 TGFSDVYAEALNPISIIVRAGSDFSSQPNDRHAFSTDGKWNFGQSEPGGVTTGCTVAAS 579  
Db 675 TSGGVKLVLYKNNETSASTGSRP-----WPK-----IVNGG--SSS 708  
Qy 580 ADGSRF---VW--APGDGPGPVVYAVGFGNSWAASQGVPAQAIRSDRVNPKTFYALSN 634  
Db 709 VDLRSVKIRYWTVDGDKPOSVC-----DW-----AOLGASNVT-FNFVKLSSG 752  
Qy 635 TFYRSTDDGVTFOVAAAGLPSSGAVGYMHFAVPKE-GDLWLAASSGLYHSTNGGSSWA 693  
Db 753 V-----SGADYY-LEVGF-SSGAGLQ-----FGKOTGDIQVRFKNDWSNYNQADDWS 800  
Qy 694 ITGVSSANVGFOKSAPGSYPVAVFVGTGTYGAYRSDDCGTTWVLINDDOHQYQNWG 753  
Db 801 LQSMTN-----YGENAKVTLY-----VDG-----VLV-----WG 824  
Qy 754 QAITGDHANLRVYICTNGRGIVYDGGAPSGSPSPSVSPSASPSPSPSPSPSPSPSPSPS 813  
Db 825 Q-----BFGGA-----TPAPTSTATPT 841  
Qy 814 PPSPPSP 873  
Db 842 PPTATPT 887  
Qy 874 KVQYKNDSPAGDNQIKPGLQVNTGSSSYDLSTVTYVYFTRGDSSTLYVNCDMAAIG 933  
Db 888 KVMYANGNLSSPTNLNPKIKIENVGTGTVADLSRVKRYWYTTIDGEATQSV----- 938

Qy 934 CGNTRASFGSVNPA 947  
Db 939 -----SVASSINPA 947

## RESULT 2

US-09-640-419C-28  
; Sequence 28, Application US/09640419C  
; Patent No. 6630615  
; GENERAL INFORMATION:  
; APPLICANT: Bidney, Dennis L  
; APPLICANT: Crasta, Oswald R  
; APPLICANT: Hu, Xu  
; APPLICANT: Lu, Guihua  
; TITLE OF INVENTION: DEFENSE-RELATED SIGNALING GENES AND METHODS OF USE  
; FILE REFERENCE: 35718/193009 (5718-92)  
; CURRENT APPLICATION NUMBER: US/09/640,419C  
; CURRENT FILING DATE: 2000-08-17  
; PRIOR APPLICATION NUMBER: 60/149,656  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/206,405  
; PRIOR FILING DATE: 2000-05-23  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 1749  
; TYPE: PRT  
; ORGANISM: Synechocystis PCC6803  
US-09-640-419C-28

Query Match 5.9%; Score 301.5; DB 2; Length 1749;  
Best Local Similarity 19.0%; Pred. No. 3.5e-11;  
Matches 225; Conservative 117; Mismatches 402; Indels 439; Gaps 46;

Qy 17 VSLIAATASFAVAALGVLPITAITASPAHAAATTPYTWSNVAIGGG----- 62  
Db 23 LALLACSSFSFG---NVLAQNTIPADPTGTGTVDAAQGNQFNIGGSLSGDQNLPHSLQ 78  
Qy 63 -----GFVDGIVFNEGAPGILYVRTDIDGMYRWA 92  
Db 79 QFGLDQGLQIANFLSNPDIRNLTIRIVGDSASIIINGLIQVSGNANLFLMNPAGMIFGPN 138  
Qy 93 A-----NGRWIPLLDVWGNNGVNGVVS----- 116  
Db 139 SINVPDGVVTTGSAIGFNDQFQV-----PSNDYNALIGNPSQAFADLANPGLIINA 193  
Qy 117 -----IAADPINTNKYMAAVGMVYTNWDNDGAILRSS----- 149  
Db 194 GDLSTEGKNLTFLAGNIVNTGSLAAPGQNTVAAVPGQNRIRISQAGSLLSLEVEVSPQ 253  
Qy 150 -DQATWQITPLPKLGNMGRGMRERLAVDPND-----NIFYGAPSGKGLWRSTD 202  
Db 254 MNQGSFSLVLDLPTLLTQAGASNLDLG--LAVQPNGSVTTNGTNAVSLPLPGSVTTISGND 311  
Qy 203 SGATWSQMTNFPDVGTYIANPTDTTGYCS---DIQ-----GVWVAFD-KSSSSILGQASK 253  
Db 312 AS---GKSTNISS-GGOVAIAGDOIAVQGATVDVSGNGGGGTVRIGGDFQGLITLPNASQ 367  
Qy 254 TIFVGVAADPNPV-----FWSRDRGGA--TWQAVPGAPTGFPHKGVFDPVNHVLYIATNT 307  
Db 368 TLI-----DSNSVVKADALLTNGGTVIWWADDSTFSGNI-----AKSLMVDVTARVNTFATMGE 444  
Qy 308 GGPVDSGSGDWKFPSTSGTWTRISVPVSTDTANDYFGYSGLTIDRQHPNTIMVATQISW 367  
Db 407 GGTMGNGG---FVETSG-----AKSLMVDVTARVNTFATMGE 444  
Qy 368 W---PDTIIFRSTDDGATWTRIND-----WTSYPNRSRYVLDISAEPLWTFGVQNPVPP 420  
Db 445 WLLDPLEIIVGTTDDLLADPKLSVLATITSLDNGNVLQADQSTAVQANFADPSAP--- 502  
Qy 421 SPKLGWMDERAMIDPNSDRMLYGTGATLYA-----TNDLTKWDSGGQIHI-- 466







RESULT 8  
 US-09-033-537A-1  
 ; Sequence 1, Application US/09033537A  
 ; Patent No. 5958083  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Onishi, Masahiro  
 ; APPLICANT: Fich, Merete  
 ; APPLICANT: Toft, Annette Hanne  
 ; APPLICANT: Sh lein, Martin  
 ; TITLE OF INVENTION: Prevention Of Back-Staining  
 ; TITLE OF INVENTION: In Stone Washing  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: NY  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10174  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: FastSEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/033,537A  
 ; FILING DATE: 02-MAR-1998  
 ; CLASSIFICATION: 008  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 0993/95  
 ; FILING DATE: 08-SEP-1995  
 ; APPLICATION NUMBER: PCT/DK96/00364  
 ; FILING DATE: 03-SEP-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Green, Reza  
 ; REGISTRATION NUMBER: 38,475  
 ; REFERENCE/DOCKET NUMBER: 4492.204-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 551 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-09-033-537A-1  
 ;  
 ; Query Match 4.4%; Score 225.5; DB 1; Length 551;  
 ; Best Local Similarity 23.8%; Pred. No. 7.9e-07;  
 ; Matches 114; Conservative 72; Mismatches 166; Indels 127; Gaps 22;  
 ;





```
Matches    37; Conservative   35; Mismatches   40; Indels     9; Gaps      3;
```

Qy 783 APGSGSPSVPSA-----SFLSPSPSSPSPSPSSPSSPSSPSPSPSPSPSRSPS 839  
      |||:::|||                |||:::|||::|||::|||::|||:  
Db 5991 APAAPAPAAPTDEAPAPAKPAKPAKPAPAKPAKPAPAKPAKPAPAPAPA 6050  
Qy 840 PSASPSPSSPSPS---SPSSSPSTTPSSPVSGGVKVQ----YKNDSAPGDNQIKPL 893  
      |||:::|||::|||            |||::|||::|||::|||:  
Db 6051 PAPKPAAPAPAPAPAKPEKPAEKPAKPAPETPKTKGWKGOMGLKEIDESDEDVVKGF 6110

Qy .894 Q 894  
Db 6111 R 6111

RESULT 13  
US-09-547-693-233  
; Sequence 233, Application US/09547693  
; Patent No. 6639050  
; GENERAL INFORMATION:  
; APPLICANT: Kieliszewski, Marcia  
; TITLE OF INVENTION: Synthetic Genes for Plant Gums and Other Hydroxyproline-Rich  
; FILE REFERENCE: Glycoproteins  
; CURRENT APPLICATION NUMBER: OHU-04089  
; CURRENT FILING DATE: US/09/547,693  
; CURRENT FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 236  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 233  
; LENGTH: 76  
; TYPE: PRT  
; ORGANISM: Artificial/Unknown  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Synthetic  
US-09-547-693-233

Query Match                     4.2%; Score 215.5; DB 2; Length 76;  
Best Local Similarity   51.8%; Pred.No. 2.7e-07;  
Matches   44; Conservative   2; Mismatches   28; Indels   11; Gaps   1;

Qy 781 GGAGSGSPSVSPSNASLSLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 840  
Db 1 GSSTRASPP-----ESP 49  
Qy 841 SASPSPSSPSPSSP 865  
Db 50 PRSP 74

RESULT 14  
US-09-328-352-5503  
; Sequence 5503, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTERIA  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: GTC99-03PA  
; CURRENT FILING DATE: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5503  
; LENGTH: 3892  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5503

Query Match                     4.1%; Score 213; DB 2; Length 3892;  
Best Local Similarity   20.7%; Pred.No. 6.9e-05;  
Matches   238; Conservative   137; Mismatches   429; Indels   348; Gaps   58;

Qy 33 GVLPITAITSPAHAAT-TOPYTWSNVAIIGGGFVDG--IVFNEG---AFGIILYVRTDIGG 86

[illegible]

|                       |              |                    |                    |                                     |
|-----------------------|--------------|--------------------|--------------------|-------------------------------------|
| Query Match           | 4.1%;        | Score 210.5;       | DB 1;              | Length 521;                         |
| Best Local Similarity | 22.2%;       | Pred. No. 7.2e-06; |                    |                                     |
| Matches 137;          | Conservative | 74;                | Mismatches 198;    | Indels 207;                         |
| Gaps                  | 31;          |                    |                    |                                     |
| QY                    | 379          | GGATWTRIWDWTSV     | PNRSRLRVLDISAE     | PMLTFGQVPNP                         |
|                       |              |                    |                    |                                     |
|                       |              | -----HTSGREILDAN   | -----NVPRIAGINWFG  | PETCNVYVHG                          |
| DB                    | 2            | GGGYW              |                    | 40                                  |
| QY                    | 428          | -----DEMAIDPNS     | DRMLYGTGATLYAT     | NDLTKWDSGGQTHIAPM                   |
|                       |              |                    |                    |                                     |
|                       |              | -----VKGLETA       | VND                | 480                                 |
| DB                    | 41           | LWSRDYKSLMDQIKS    | -----LGYNITIRLPYS  | DDILKPGTWPNSINFYQMNODLOGLTSLQVMD    |
|                       |              |                    |                    |                                     |
|                       |              | -----98            |                    |                                     |
| QY                    | 481          | LISPPSGAPLISALGDUG | -----GFTHADVTAVP   | STIFTSPVFTTGTSDVDAEL                |
|                       |              |                    |                    |                                     |
|                       |              | -----529           |                    |                                     |
| DB                    | 99           | KI                 | -----VAYAGOIGLRIL  | DRHPDCSGQSALWYTSVSEATWISDLQALAQRYKG |
|                       |              |                    |                    |                                     |
|                       |              | -----150           |                    |                                     |
| QY                    | 530          | NPSIIVPAGSFDPSS    | QPNDRHVAESTDGGK    | NWFGSGEPGG                          |
|                       |              |                    |                    |                                     |
|                       |              | -----VTTGCTVAA     |                    | 578                                 |
| DB                    | 151          | NPTVV              | -----GFDLHNEPHDPAC | WCGGDPISDWRLLAAERAGNAVL             |
|                       |              |                    |                    |                                     |
|                       |              | -----VFEBG         |                    | -VQ 204                             |
| QY                    | 579          | SADGSREVAAPGDPG    | QGPVAVVAVFGNSWAA   | SQ-----GVP-----ANAOIRSDRVNPKTFY     |
|                       |              |                    |                    |                                     |
|                       |              | -----629           |                    |                                     |

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:20:38 ; Search time 184.353 Seconds  
(without alignments)  
2169.009 Million cell updates/sec

Title: US-09-917-376-1  
Perfect score: 5134  
Sequence: 1 MDRSENRLTWSRRLVSL.....RSPGSVNPATPTADTVLQX 957

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pcp:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pcp:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pcp:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pcp:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pcp:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID | Description         |
|------------|--------|-------------|--------|----|---------------------|
| 1          | 5134   | 100.0       | 957    | 3  | US-09-917-376-1     |
| 2          | 5134   | 100.0       | 957    | 4  | US-10-155-400-1     |
| 3          | 4036   | 78.6        | 740    | 3  | US-09-917-376-3     |
| 4          | 4036   | 78.6        | 740    | 3  | US-09-917-376-6     |
| 5          | 4036   | 78.6        | 740    | 4  | US-10-155-400-3     |
| 6          | 4036   | 78.6        | 740    | 4  | US-10-155-400-6     |
| 7          | 2478   | 48.3        | 882    | 4  | US-10-156-761-9395  |
| 8          | 1680   | 32.7        | 726    | 3  | US-09-917-376-7     |
| 9          | 1680   | 32.7        | 726    | 4  | US-10-155-400-7     |
| 10         | 1625.5 | 31.7        | 838    | 4  | US-10-420-191-2     |
| 11         | 1604   | 31.2        | 818    | 4  | US-10-026-994-2     |
| 12         | 1442   | 28.1        | 739    | 4  | US-10-156-761-10111 |
| 13         | 1125   | 21.9        | 812    | 4  | US-10-395-241-12    |
| 14         | 1117   | 21.8        | 789    | 4  | US-10-395-241-14    |
| 15         | 1114   | 21.7        | 826    | 4  | US-10-395-241-18    |
| 16         | 848    | 16.5        | 555    | 3  | US-09-927-827-47    |
| 17         | 823    | 16.0        | 1228   | 3  | US-09-917-384-1     |
| 18         | 823    | 16.0        | 1228   | 3  | US-09-917-383-1     |
| 19         | 702    | 13.7        | 762    | 3  | US-09-917-378-1     |
| 20         | 511.5  | 10.0        | 1043   | 3  | US-09-917-384-6     |
| 21         | 511.5  | 10.0        | 1043   | 3  | US-09-917-383-6     |
| 22         | 462    | 9.0         | 88     | 3  | US-09-917-376-5     |
| 23         | 462    | 9.0         | 88     | 4  | US-10-155-400-5     |
| 24         | 462    | 9.0         | 89     | 3  | US-09-917-376-4     |
| 25         | 462    | 9.0         | 89     | 4  | US-10-155-400-4     |
| 26         | 462    | 9.0         | 154    | 3  | US-09-917-378-4     |
| 27         | 459    | 8.9         | 150    | 3  | US-09-917-384-5     |

|    |       |     |      |   |                      |                   |
|----|-------|-----|------|---|----------------------|-------------------|
| 28 | 459   | 8.9 | 150  | 3 | US-09-917-383-5      | Sequence 5, Appli |
| 29 | 389   | 7.6 | 284  | 5 | US-10-418-032-270    | Sequence 270, App |
| 30 | 346.5 | 6.7 | 901  | 5 | US-10-939-262-12     | Sequence 12, Appl |
| 31 | 336.5 | 6.6 | 741  | 4 | US-10-156-761-8100   | Sequence 8100, Ap |
| 32 | 326.5 | 6.4 | 628  | 5 | US-10-939-262-14     | Sequence 14, Appl |
| 33 | 288   | 5.6 | 96   | 4 | US-10-437-708-236    | Sequence 236, App |
| 34 | 288   | 5.6 | 96   | 4 | US-10-395-402-112    | Sequence 112, App |
| 35 | 288   | 5.6 | 96   | 5 | US-10-257-193-236    | Sequence 236, App |
| 36 | 288   | 5.6 | 96   | 5 | US-10-418-032-236    | Sequence 236, App |
| 37 | 274   | 5.3 | 406  | 5 | US-10-450-763-57609  | Sequence 57609, A |
| 38 | 268   | 5.2 | 284  | 5 | US-10-418-032-271    | Sequence 271, App |
| 39 | 262   | 5.1 | 2468 | 4 | US-10-246-330-4      | Sequence 4, Appli |
| 40 | 262   | 5.1 | 2468 | 4 | US-10-282-122A-66335 | Sequence 66335, A |
| 41 | 261.5 | 5.1 | 599  | 3 | US-09-955-555A-29    | Sequence 29, Appl |
| 42 | 253   | 4.9 | 2435 | 4 | US-10-282-122A-47453 | Sequence 47453, A |
| 43 | 252.5 | 4.9 | 2117 | 4 | US-10-120-801-63     | Sequence 63, Appl |
| 44 | 246   | 4.8 | 1066 | 4 | US-10-647-196-43     | Sequence 43, Appl |
| 45 | 243.5 | 4.7 | 1049 | 4 | US-10-282-122A-49900 | Sequence 49900, A |

ALIGNMENTS

RESULT 1

US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

|                       |                 |                                  |                               |             |
|-----------------------|-----------------|----------------------------------|-------------------------------|-------------|
| Query Match           | 100.0%          | Score 5134;                      | DB 3;                         | Length 957; |
| Best Local Similarity | 100.0%          | Pred. No. 3.9e-280;              | Mismatches 0;                 | Indels 0;   |
| Matches 956;          | Conservative 0; |                                  |                               | Gaps 0;     |
| Qy                    | 1               | MDRSENRLTWSRRLVSLAATASFAVAALGVLP | IAITASPAHAATTPYTSNVAIG        | 60          |
| Db                    | 1               | MDRSENRLTWSRRLVSLAATASFAVAALGVLP | IAITASPAHAATTPYTSNVAIG        | 60          |
| Qy                    | 61              | GGGFVDGIVFNEGAGPILYVRTDIGMWRDAA  | ANGRWIPLLDWGNNNGYNGVSIAD      | 120         |
| Db                    | 61              | GGGFVDGIVFNEGAGPILYVRTDIGMWRDAA  | ANGRWIPLLDWGNNNGYNGVSIAD      | 120         |
| Qy                    | 121             | PINTNKWAAVGMVYNSWDPNDGAILRSSDQAT | WQITLPEKLGNNMGRGNGERLAVD      | 180         |
| Db                    | 121             | PINTNKWAAVGMVYNSWDPNDGAILRSSDQAT | WQITLPEKLGNNMGRGNGERLAVD      | 180         |
| Qy                    | 181             | PNNNLYFGAPSGKGLWRSTDSGATWSQMTN   | FPDVGTYIANPTDTTGYQSDIQGVVWA   | 240         |
| Db                    | 181             | PNNNLYFGAPSGKGLWRSTDSGATWSQMTN   | FPDVGTYIANPTDTTGYQSDIQGVVWA   | 240         |
| Qy                    | 241             | FDKSSSSLGQASKTIFVGVDADPNNFVWSR   | DGGATWQAVPGAPTGFIPHKGVDPVNVHL | 300         |
| Db                    | 241             | FDKSSSSLGQASKTIFVGVDADPNNFVWSR   | DGGATWQAVPGAPTGFIPHKGVDPVNVHL | 300         |





```

; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Catalytic domain GH74
US-09-917-376-3

```

```

Query Match      78.6%; Score 4036; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.5e-218;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
DB 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 107 NNWYGVVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQGTATQITPLPFLKGG 166
DB 61 NNWYGVVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQGTATQITPLPFLKGG 120
QY 167 NMPGRGMRGLAVDPNNNDNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 226
DB 121 NMPGRGMRGLAVDPNNNDNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWAFDKSSSLGQASKTIIFGVADPNNPVFWSRDGGATWQAVPGATGF 286
DB 181 TGYQSDIQGVVWAFDKSSSLGQASKTIIFGVADPNNPVFWSRDGGATWQAVPGATGF 240
QY 287 IPHKGVPDPVNHVLYIATNTGGPYDGGSDGVKFSVTSGTWTTRI SPVPSDTTANDYFGY 346
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDGGSDGVKFSVTSGTWTTRI SPVPSDTTANDYFGY 300
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPNRSRLRYVLDISAE 406
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPNRSRLRYVLDISAE 360
QY 407 PWLTFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
DB 361 PWLTFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGSVDY 526
DB 421 APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGSVDY 480
QY 527 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGCTVAASADGSRFV 586
DB 481 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGCTVAASADGSRFV 540
QY 587 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
QY 647 QPVAAGLPSSGAVGVNMFHAPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706
DB 601 QPVAAGLPSSGAVGVNMFHAPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPGSSYPAVFVVGTTGGVTVGAYRSDCCGTTWVLINDDHOYGNWGOAITGDHANLRRV 766
DB 661 KSAPGSSYPAVFVVGTTGGVTVGAYRSDCCGTTWVLINDDHOYGNWGOAITGDHANLRRV 720
QY 767 YITNGRGIVYDGGAPSG 786
DB 721 YITNGRGIVYDGGAPSG 740

```

```

RESULT 4
US-09-917-376-6
; Sequence 6, Application US/09917376
; Publication No. US20040038334A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 740
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; US-09-917-376-6

```

```

Query Match      78.6%; Score 4036; DB 3; Length 740;
Best Local Similarity 100.0%; Pred. No. 1.5e-218;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106
DB 1 ATTQPYTWSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
QY 107 NNWYGVVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQGTATQITPLPFLKGG 166
DB 61 NNWYGVVSIADPINTNKWAAVGMVYNSWDNDGAILRSSDQGTATQITPLPFLKGG 120
QY 167 NMPGRGMRGLAVDPNNNDNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 226
DB 121 NMPGRGMRGLAVDPNNNDNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180
QY 227 TGYQSDIQGVVWAFDKSSSLGQASKTIIFGVADPNNPVFWSRDGGATWQAVPGATGF 286
DB 181 TGYQSDIQGVVWAFDKSSSLGQASKTIIFGVADPNNPVFWSRDGGATWQAVPGATGF 240
QY 287 IPHKGVPDPVNHVLYIATNTGGPYDGGSDGVKFSVTSGTWTTRI SPVPSDTTANDYFGY 346
DB 241 IPHKGVPDPVNHVLYIATNTGGPYDGGSDGVKFSVTSGTWTTRI SPVPSDTTANDYFGY 300
QY 347 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPNRSRLRYVLDISAE 406
DB 301 SGLTIDRQHPNTIMVATQISWMPDTIIFRSTDGATWTRIMDWTSPNRSRLRYVLDISAE 360
QY 407 PWLTFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 466
DB 361 PWLTFGVQPNPPVPSPKLGWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420
QY 467 APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGSVDY 526
DB 421 APMVKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIETSPVFTTGSVDY 480
QY 527 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGCTVAASADGSRFV 586
DB 481 AELNPSIIVRAGSDPPSQPNDRHVAFTDGGKWNFGQSEPGVTTGCTVAASADGSRFV 540
QY 587 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFYRSTDGGVTF 646
DB 541 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600
QY 647 QPVAAGLPSSGAVGVNMFHAPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706
DB 601 QPVAAGLPSSGAVGVNMFHAPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
QY 707 KSAPGSSYPAVFVVGTTGGVTVGAYRSDCCGTTWVLINDDHOYGNWGOAITGDHANLRRV 766

```

Db 661 KSAFGSSYPVAVVGTGGTGAIRSDCCGTTWVLINDDQHOYGNWQAITGDHANLRRV 720  
Qy 767 YIGTNGRGIVYDGGAPSG 786  
Db 721 YIGTNGRGIVYDGGAPSG 740  
RESULT 5  
US-10-155-400-3  
; Sequence 3, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-10-155-400-3  
Query Match 78.6%; Score 4036; DB 4; Length 740;  
Best Local Similarity 100.0%; Pred. No. 1.5e-218;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 47 ATTQPTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 106  
Db 1 ATTQPTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
Qy 107 NMGVNGVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 166  
Db 61 NMGVNGVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 120  
Qy 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWTRIDWNTSPNRSRLRVLDISAE 226  
Db 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWTRIDWNTSPNRSRLRVLDISAE 180  
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNVFWSRDGGATWQAVPGAPTGF 286  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNVFWSRDGGATWQAVPGAPTGF 240  
Qy 287 IPHKGVPDPVNHVLIATSNVTGPGYDSSGDVWKFVSTGWTWTRISVPVSTDTANDYFGY 346  
Db 241 IPHKGVPDPVNHVLIATSNVTGPGYDSSGDVWKFVSTGWTWTRISVPVSTDTANDYFGY 300  
Qy 347 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWTRIDWNTSPNRSRLRVLDISAE 406  
Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWTRIDWNTSPNRSRLRVLDISAE 360  
Qy 407 PWLTFGQVQNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 466  
Db 361 PWLTFGQVQNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Qy 467 APVKGLEETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVTTGTSVDY 526  
Db 421 APVKGLEETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVTTGTSVDY 480  
Qy 527 AELNPSIIVRAGSFDSSQNDRHVAFSTDDGKNWFQSGEPGGVTTGGTVAASADSRFV 586  
Db 481 AELNPSIIVRAGSFDSSQNDRHVAFSTDDGKNWFQSGEPGGVTTGGTVAASADSRFV 540  
Qy 587 WAPGDPGPVYAVFGNSWAASQGVPAQAIRSDRNPKTFFYALNSNGTFYRSTDDGVT 646

Db 541 WAPGDPGPVYAVFGNSWAASQGVPAQAIRSDRNPKTFFYALNSNGTFYRSTDDGVT 600  
Qy 647 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706  
Db 601 QPVAAGLPSSGAVGMFHAVPKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
Qy 707 KSAFGSSYPVAVVGTGGTGAIRSDCCGTTWVLINDDQHOYGNWQAITGDHANLRRV 766  
Db 661 KSAFGSSYPVAVVGTGGTGAIRSDCCGTTWVLINDDQHOYGNWQAITGDHANLRRV 720  
Qy 767 YIGTNGRGIVYDGGAPSG 786  
Db 721 YIGTNGRGIVYDGGAPSG 740  
RESULT 6  
US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-10-155-400-6  
Query Match 78.6%; Score 4036; DB 4; Length 740;  
Best Local Similarity 100.0%; Pred. No. 1.5e-218;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 47 ATTQPTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 106  
Db 1 ATTQPTWSNVAIGGGGFVDGIVFNEGAPGILYVRTDIGMYRWDANGRWIPLLDWVGW 60  
Qy 107 NMGVNGVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 166  
Db 61 NMGVNGVSIADPINTNKVAAVGMVYTNWDNDGAILRSSDQATWQITPLPKLGG 120  
Qy 167 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWQITPLPKLGG 226  
Db 121 NMPGRGGERLAVDPNNNDILYFGAPSGKGLWRSTDGATWQITPLPKLGG 180  
Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNVFWSRDGGATWQAVPGAPTGF 286  
Db 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNVFWSRDGGATWQAVPGAPTGF 240  
Qy 287 IPHKGVPDPVNHVLIATSNVTGPGYDSSGDVWKFVSTGWTWTRISVPVSTDTANDYFGY 346  
Db 241 IPHKGVPDPVNHVLIATSNVTGPGYDSSGDVWKFVSTGWTWTRISVPVSTDTANDYFGY 300  
Qy 347 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWTRIDWNTSPNRSRLRVLDISAE 406  
Db 301 SGLTIDRQHPNTIMVATQISWNPDTIIFRSTDDGATWTRIDWNTSPNRSRLRVLDISAE 360  
Qy 407 PWLTFGQVQNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 466  
Db 361 PWLTFGQVQNPVPSPKLGWDEAMADPNSDRMLYGTGATLYATNDLTWDSGGQIHI 420  
Qy 467 APVKGLEETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVTTGTSVDY 526  
Db 421 APVKGLEETAVNDLISPPSGAPLISALDGLGGFTHADVTAVPSTIFTSPVTTGTSVDY 480

QY 527 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPCGVTTGGTVAASADGSRPV 586  
DB 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPCGVTTGGTVAASADGSRPV 540  
QY 587 WAPGDPGPVVYAVGFGNSWAASQGVPAANOIRSDRVNPKTFYALNSGTFFRSTDDGGVTF 646  
DB 541 WAPGDPGPVVYAVGFGNSWAASQGVPAANOIRSDRVNPKTFYALNSGTFFRSTDDGGVTF 600  
QY 647 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSSAITGVSSAVNVGFG 706  
DB 601 QPVAAGLPSSGAVGMFHAVPGKEGDLMLAASSGLYHSTNGSSSAITGVSSAVNVGFG 660  
QY 707 KAPGSSYPVAVVGTIGVTCAYRSDCGTTWVLINDQHQYGNWGOAITGDHANLRV 766  
DB 661 KAPGSSYPVAVVGTIGVTCAYRSDCGTTWVLINDQHQYGNWGOAITGDHANLRV 720  
QY 767 YIGTNGRGIVYGDIGGAPSG 786  
DB 721 YIGTNGRGIVYGDIGGAPSG 740

## RESULT 7

US-10-156-761-9395

; Sequence 9395, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9395  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis

US-10-156-761-9395

Query Match 48.3%; Score 2478; DB 4; Length 882;

Best Local Similarity 50.4%; Pred. No. 7.1e-131; Mismatches 227; Indels 110; Gaps 14; Matches 473; Conservative 128;

QY 23 TASFAVAAL--GVLPIAITASPAHAATQPYTWSNVAIGGGFVDGIVFNEGAPGILYV 80  
DB 8 TAVLALAGLPAGTPPALAASAPTATIAATYSKNAARVDGCGFVGVFVFNSEKNLAYA 67  
QY 81 RTDIGMTRWDAANGRWITPLLDWGNWNGVVSIAADPINTNKVMAAGVMTNSWDP 140  
DB 68 RTDIGAYRWAESKWTWPLDLSVGWSWGHTGVVSLASDSVDPNKVAAVGTYNWDP 127  
QY 141 NDGATLRSSDQATWQITPLPKLGNMPPGMRGERLAVDPNNDILYFAPSGKGLWRS 200  
DB 128 GNGAVIRSGDRGASQKQKTDLPKLGNNPPGMRGERLAVDPNRSVLYIGAPSGKGLWRS 187  
QY 201 TDSGATWSQMTNFPDVGTYIANPTDTTGTQSDIQGVVWVAFDKSSSLGQASKTIFVGVA 260  
DB 188 TDSGASNSQVTDFFNVGTYVQDADTSGASDNQGIWMTFDESGSGSRTTIVGVA 247  
QY 261 DPNPNFVWRDGGATWQAVPGATGFIPIHKGVFDPVNNHVLXIATSNTPGPDGSSGVNWK 320  
DB 248 DXDNSVYRSTAGATWSRLAGQPTGHLAKHGVLDAAANGCLYLAYS DKGPGYDGGKGLWR 307

QY 321 FSVTSGTWRIISVPSTDTANDYFGYSGLTIDROHPNTIMVATQISWMPDIIIFRSTDDG 380  
DB 308 YTTKGTWTNINISFVAEADT---YFGESGLTVDRQHFGTVMATAYSWMPDTQLFRSTDSG 364  
QY 381 AWTIRIDWTSVPNSRLRYVLDISAEPLWLTFGQPNPVPSPKLGWMADEAMADPNSDR 440  
DB 365 GTWTXANDYTSVPNSRNRFTMDVSSPWLWTGANPAPPEQTPKLGWMTESLEIDPFDSAR 424  
QY 441 MLYGTGATLYATNDLTWDSGGQIHIAPMVKGLEETAANDLISPPS--GAPLISALGDLGG 499  
DB 425 WMYGTGATVYGTNDLTNWDSSGQFTIKPMARGLEETAANDLASPPSGGAQLFSAALGDI 484  
QY 500 FTHADVAVPSTIFSPVETTTGTSVDYAEINPSIIVRAGSFDPSOPNDRHVAFTSDGGK 559  
DB 485 FRHTDLTTVPISLMTYSPNFTTSLDYAETDGTVVRVGNLD--SGP---HVAFTSDNGA 539  
QY 560 NMFQSEPGGVTTGGTVAASADGSRFVWAPGDPQPVVAVGFGNSWAASQGVPAQAIR 619  
DB 540 NMFAGADPSGVSGGTVAAASDGRFVWSPAGTG---VQYTTGFTGWSASAGLPAGAVE 597  
QY 620 SDRVNPKTFYALNSGTFFRSTDDGGVTFQPVAA--GLPSSGAVGMFHAVPGKEGDLMLAAS 678  
DB 598 SDRVNPKTFYGFSGRFFVYSSDGGATFTASAATGLPSGDS--VRFKALPCTKGDIMLAGG 655  
QY 679 S-----GLYHSTNGSSSAITGVSSAVNVGFGKSAPGSYPVAVVGTIGVTCAYRSD 733  
DB 656 ASDGAYGLWHSTDDGGAFTKLATVDQADTIGFKAATGASQYTLTYSAKIGGVGRGIFRST 715  
QY 734 DCGTTWVLINDQHQYGNWGOAITGDHANLRVYIGTNGRGIVYGDIGGAPSGSPSPSVS 793  
DB 716 DKGASWTRVNDDAHOMGWTGAATGDPVYGRVYVSTNGRGIVYGTAGSSDGG----- 769  
QY 794 PSASPSLSPSPPSSSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 853  
DB 770 ----- 769  
QY 854 SSPSSSP 913  
DB 770 ---GTEPAPT-----GACTVYRITNOMSGGFQ--ADVQLANTGCTAWDWSLG---W 814  
QY 914 FTRDGSSTLVNCDMAAIGCG-----NTRASFGS 943  
DB 815 SPGDGOEVTLQWNASYAQAGSGVTAAANLAWNGRVAAGS 852

RESULT 8  
US-09-917-376-7  
; Sequence 7, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Aspergillus aculeatus  
US-09-917-376-7

Query Match 32.7%; Score 1680; DB 3; Length 726;  
Best Local Similarity 46.2%; Pred. No. 4e-86;  
Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;

QY 47 ATTOPVYTSNVAI--GGGGFVDGIVFNEGAPGILYVRTDIGMTRWDAANGRWITPLLDWVG 105

Db 1 AASQAYTWKVVVTTGGGGFTPGIVFNPSAKGVAVARTDGGAYRLN-SDDTWTPMLMDWVG 59  
 Qy 106 ---WNNWYNGVGIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPF 162  
 Db 60 NDTWHDW---GIDALATDPVTDTRVYVAVGMVYNTWEDPNVGSILRSTDDGDTWTETKLPF 116  
 Qy 163 KLGNNMPGRGGERLAVDPNNNDILYFCAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 222  
 Db 117 KVGNNMPGRGGERLAVDPNNKNSILYFCARSGHGLWKSTDYCATWSNVTSTFTWTGYFQD 176  
 Qy 223 PTDTTGYQSDIQGVVWVAFKSSSSLSGOASKTIIFGVADPNPNPFWRSRGGATWQAVPGA 282  
 Db 177 SSST--YTSDPVGAIAWTFDSTSGSSGATPRIFGVADAGKSVFKSEADAGATWAVSGE 234  
 Qy 283 PT-GFIPHKGVDPVNVHLVYATISNTGPGYDSSGDVWKFVSTSGTWTTRISPVPSRSLRYVL 401  
 Db 292 TYGYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIWENMGYPNSIYYYSV 351  
 Qy 402 DISAEPWLTFTGQVQNP-PVPSFKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDS 460  
 Db 352 DISNAPWIDTSTTDQFV---RVGMVVEALAIIDPFNSHMLYGTGLTVYGGHDLTNWDS 408  
 Qy 461 GGOIHIAPMVKGLEETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTT 520  
 Db 409 KHNVTVKSIAVGIEMAVLGLITPPGPAALLSAGVDDGGFVHSDLDAAAPNQAYHTPYGT 468  
 Qy 521 GTSVDYAEIAPNSIIVRAGSFPDSSQPNDRHVAFTDGGKNNFQSGEPGVTTCGTVAASA 580  
 Db 469 TNGIDYAGNKPNSIVRSGASD--DYPT---LALSNNFGSTWYADYAASTSTGTGAVALSA 523  
 Qy 581 DGSFVWAPGDPQPVVYAVFGNSWAASQGVANAIQIRSDRVNPKTFYALNSGTFYRST 640  
 Db 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSSLPSGAVIASDKSNTVFGGSAGAIYVSK 580  
 Qy 641 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAI-TGVSS 699  
 Db 581 NTATSFYKTVS-LGSSITVNAI-RAHPSIAGDVWASTDKGLWHSDYGSTFTQIGSGVTA 638  
 Qy 700 ANVFGFKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDQHOYGNWGOA--IT 757  
 Db 639 GWSFGFKASSTGSYVVIYGFFTIDGAAGLKFSEADAGTNQVSDASHGFGS-GSANVNV 697  
 Qy 758 GDHANLRVYIGTNGRGIVYDGGAPSG 786  
 Db 698 GDLQYGRVFRGHRPGLLRQSQREPAG 726

RESULT 9

US-10-155-400-7  
 ; Sequence 7, Application US/10155400  
 ; Publication No. US20030108989A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYLICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 726  
 ; TYPE: PRN  
 ; ORGANISM: Aspergillus aculeatus  
 ; US-10-155-400-7

Query Match 32.7%; Score 1680; DB 4; Length 726;  
 Best Local Similarity 46.2%; Fred. No. 4e-86;  
 Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;  
 Qy 47 ATTQBYTWSNVAI---GGGGFVDGIVFNEGAPGILYVYRTDIDGMYRMDAANGRWIPLLDWVG 105  
 Db 1 AASQAYTWKVVVTTGGGGFTPGIVFNPSAKGVAVARTDGGAYRLN-SDDTWTPMLMDWVG 59  
 Qy 106 ---WNNWYNGVGIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPF 162  
 Db 60 NDTWHDW---GIDALATDPVTDTRVYVAVGMVYNTWEDPNVGSILRSTDDGDTWTETKLPF 116  
 Qy 163 KLGNNMPGRGGERLAVDPNNNDILYFCAPSGKGLWRSTDSGATWSQMTNFPDVGTYIAN 222  
 Db 117 KVGNNMPGRGGERLAVDPNNKNSILYFCARSGHGLWKSTDYCATWSNVTSTFTWTGYFQD 176  
 Qy 223 PTDTTGYQSDIQGVVWVAFKSSSSLSGOASKTIIFGVADPNPNPFWRSRGGATWQAVPGA 282  
 Db 177 SSST--YTSDPVGAIAWTFDSTSGSSGATPRIFGVADAGKSVFKSEADAGATWAVSGE 234  
 Qy 283 PT-GFIPHKGVDPVNVHLVYATISNTGPGYDSSGDVWKFVSTSGTWTTRISPVPSRSLRYVL 401  
 Db 292 TYGYGGLSVLDQVPGTLMVAALNCWPDDELIFRSTDSGATWSPIWENMGYPNSIYYYSV 351  
 Qy 402 DISAEPWLTFTGQVQNP-PVPSFKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTWKDS 460  
 Db 352 DISNAPWIDTSTTDQFV---RVGMVVEALAIIDPFNSHMLYGTGLTVYGGHDLTNWDS 408  
 Qy 461 GGOIHIAPMVKGLEETAVNDLISPPSGAPLISALDGLGFTHADVTAVPSTIFTSPVFTT 520  
 Db 409 KHNVTVKSIAVGIEMAVLGLITPPGPAALLSAGVDDGGFVHSDLDAAAPNQAYHTPYGT 468  
 Qy 521 GTSVDYAEIAPNSIIVRAGSFPDSSQPNDRHVAFTDGGKNNFQSGEPGVTTCGTVAASA 580  
 Db 469 TNGIDYAGNKPNSIVRSGASD--DYPT---LALSNNFGSTWYADYAASTSTGTGAVALSA 523  
 Qy 581 DGSFVWAPGDPQPVVYAVFGNSWAASQGVANAIQIRSDRVNPKTFYALNSGTFYRST 640  
 Db 524 DGDVTLLMSSTSGALVSKSQG---TLTAVSSLPSGAVIASDKSNTVFGGSAGAIYVSK 580  
 Qy 641 DGGVTFQVAAAGLPSGAGVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAI-TGVSS 699  
 Db 581 NTATSFYKTVS-LGSSITVNAI-RAHPSIAGDVWASTDKGLWHSDYGSTFTQIGSGVTA 638  
 Qy 700 ANVFGFKSAPGSSYPVAVFVGTIGVGTAYRSDDCGTTWVLINDDQHOYGNWGOA--IT 757  
 Db 639 GWSFGFKASSTGSYVVIYGFFTIDGAAGLKFSEADAGTNQVSDASHGFGS-GSANVNV 697  
 Qy 758 GDHANLRVYIGTNGRGIVYDGGAPSG 786  
 Db 698 GDLQYGRVFRGHRPGLLRQSQREPAG 726

RESULT 10

US-10-420-191-2  
 ; Sequence 2, Application US/10420191  
 ; Publication No. US20040067569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
 ; APPLICANT: Rey, Michael W.  
 ; APPLICANT: Zaretsky, Elizabeth J.  
 ; APPLICANT: Haas, Jeffrey A.  
 ; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
 ; TITLE OF INVENTION: Encoding Same  
 ; FILE REFERENCE: 10210.200-US  
 ; CURRENT APPLICATION NUMBER: US/10/420,191  
 ; CURRENT FILING DATE: 2003-04-18  
 ; PRIOR APPLICATION NUMBER: US 60/373,987  
 ; PRIOR FILING DATE: 2002-04-19

```
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-420-191-2

Query Match      31.7%; Score 1625.5; DB 4; Length 838;
Best Local Similarity 41.7%; Pred. No. 5.5e-83;
Matches 350; Conservative 131; Mismatches 296; Indels 63; Gaps 23;

Qy 25 SFVAALGVLPALITASPAAHAATQPYTWGNVAI-GGGGFVDGIVFNEGAPGILYVTD 83
Db 4 SRVLALVLGAV-----IPAHAA-----PSWKVKLGCGGFGVPGIIFHPKTKGVAYARTD 53

Qy 84 IGGMYRMDAANGRIPLLDWV-----GNNWYGVGVVSTAAADPINTNKVWAAVGMVYNSWD 139
Db 54 IGGLYRLUN-ADDSMTAVTDGADNAGHNW---GIDAVALDPQDDQKQVYAAVGMVYNSWD 109

Qy 140 PNDGAILRSSDQGATWQITPLPFKLGNNPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWR 199
Db 110 PSNGAILRSSDQATWSTFNLPFKVGNMPCRGAGERLAVDPANSNIIYFGARSGNGLWK 169

Qy 200 STDGATWSQMTNFPDVGTYIANPTDTTGYQSDIQGVVWVAFDKSSSLGQASKTIFVGV 259
Db 170 STDGVTFSKVSSTATCTYIPDPSDSNGYNSDKQLMWVTFDSTSTTGCATSRIFVGT 229

Qy 260 ADP-NNPVFWSRDGATWQAVPGAPTGFIPHKGVDPDNNHLYIATNTGPGPYGSSGDV 318
Db 230 ADNITASVYVSTNAGTWSAVPGQPGKYPFKAKLQPAEKALYLYSDGTGYDGLGSV 289

Qy 319 WKFSVTSWTWIRISVPSTDTANDYFGYSGLTIDROHENTIMVATQISWPDITLIFRSTD 378
Db 290 WRIDAGTWDKDIIPVSGSDL---YFGGGLDLQKPGTLVVASLNSWPDQALFRSTD 346

Qy 379 GGATWTRIDWTSYPNRSLRYLDISAPFWLTFG-VQPNPPVPS-----PKLGWMDAMAI 433
Db 347 SGTWSPINAWASYPTETYYISITPPKAPWKNNFIDVTSSEPSDGLIKRLGWMIESLEI 406

Qy 434 DPFNSDRMLYGTGATLYATNDLTWKDSGGQTHIAPMKVGLBETAVNDLISPPSGAPLISA 493
Db 407 DPTDSNHLVYGTGMTIFGCHDLTNWDTNRHNSIQSLADGIEEFVSQDLASAPGSELLAA 466

Qy 494 LGDLGGFTHA---DVTAVPSTIFTSPTTCTSDVYAEINPSIIVRAGSPDSSQPNDRH 550
Db 467 VGDNGFTFASRNDLGTSPQVWATPTWATSTSDVYAGNSVKSVVRVGNVNTAGTQQ----- 521

Qy 551 VAFSTDGKNWFQSGEPGVTTGGTVAASADGRFVWAPGPGQPVVYAVFGNSWAAASQ 610
Db 522 VAISDGGATWSIDYAADTSMNGGTVAASADGDTILWSTASSG---VQRSQFGSFAVS 578

Qy 611 GVPANAQIRSRVNPKPTFYALNSGTIFYRSTDPGVTFPQVAAGLPSSGAVGM---FHAVPG 668
Db 579 SLPAGAVIASDKTNSVYAGSGSTFYVSKDTGSSF---TRG-PKLSAGTIRIDIAAHT 634

Qy 669 KEGDLWLAASGLYHSTNGSSWAI-TGVSSAVNVGFKSAPGSSYPAVFVGTIGCVT 727
Db 635 TAGTLYVSTDVGIIFRSTDSGTTFQGVSTALTNYQIALGVGS-GSNW-NLYAFGT--GPS 690

Qy 728 GA--YRSDDCGTWVLINDDOHQYGNWQOALTGDHANLRRVYIGTNGRGIYV---GDIGGA 783
Db 691 GARLYASDGSASMTDIOGSGFGSISDTKVAGSGGTAGQVYVGTNGRGVFAQGTGVGG 750

Qy 784 PSGSPSPSVSPASPLSPSPSPSSSPSPSPSSSPSPSPSPSPSPSPSPSPSPSPSPSPSP 843
Db 751 TGGT-----SSSTKQSSSTSSASSTTLSSSVVSTRASTVTSRSTSSAAGTGS 801

; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: BGVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-026-994-2

Query Match      31.2%; Score 1604; DB 4; Length 818;
Best Local Similarity 41.8%; Pred. No. 8.7e-82;
Matches 340; Conservative 128; Mismatches 291; Indels 54; Gaps 21;

Qy 52 YTWSNVAI-GGGGFVDGIVFNEGAPGILYVTDIGMYRMDAANGRIPLLDWV-----GW 106
Db 2 PSWKVKLGCGGFGVPGIIFHPKTKGVAYARTDIGLYRLN-ADDSMTAVTDGADNAGW 60

Qy 107 NNGWYGVGVVSTAAADPINTNKVWAAVGMVYNSWDPNPDGAILRSSDQATWQITPLPKLGG 166
Db 61 HNW---GIDAVALDPQDDQKQVYAAVGMVYNSWDPSNGAILRSSDQATWSTFNLDPKVG 117

Qy 167 NMPGRGMRGLAVDPNNNDNIIYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 226
Db 118 NMPGRGAGERLAVDPANSNIIYFGARSGNGLMKSTDGVTFSKVSSTATGTYIPDPSDS 177

Qy 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIFVGVADP-NNPVFWSRDGATWQAVPGAPT 285
Db 178 NGYNSDKQLMWVTFDSTSTTGCATSRIFVGTADNITASVYVSTNAGTWSAVPGQPGK 237

Qy 286 FIPHKGVDPDNNHLYIATNTGPGYDGSQGVWKEVSTSGTWTRISVPSTDTANDYFG 345
Db 238 YPPHKAKLQPAEKALYLYSDGTGYDGLGSVWRYDIAGGTWKDITPVGSGSDL---YFG 294

Qy 346 YSGLTIDROHPNTIMVATQISWPDITLIFRSTDGATWTRIDWTSYPNRSLRYLDISA 405
Db 295 FGGLGLDLQKPGTLVVASLNSWPDQALFRSTDSGTTNSPIWAWASYPTETYYISSTPK 354

Qy 406 EPWLTFG-VQPNPPVPS-----PKLGWMDAMAIIDPFNSDRMLYGTGATLYATNDLTWKDS 460
Db 355 APWKNNFIDVTSSEPSDGLIKRLGWMIESLEIDPTDSNHLVYGTGMTIFGCHDLTNWDT 414

Qy 461 GGQTHIAPMKVGLBETAVNDLISPPSGAPLISALDGLGFTHA---DVTAVPSTIFTSVP 517
Db 415 RHNVSIQSLADGIEEFVSQDLASAPGSELLAAVGDNDGFTTFASRNDLGTSPQVWATPT 474

Qy 518 FTTGTSVDYAEINPSIIVRAGSPDSSQPNDRHAFSTGDKNWFQGSSEGGVTTGCTVA 577
Db 475 WATSTSDVYAGNSVKSVVRVGN-----TAGTQVALSSDGGATWSIDYAADTSMNGGTVA 528

Qy 578 ASADGRFVWAPGDPGQPVVYAVFGNSWAAASQGVPPANAQIRSRVNPKPTFYALNSGTFF 637
Db 529 YSADGDTILWSTASSG---VQRSQFGSFAVSLLPAGAVIASDKTNSVYAGSGSTFF 585

Qy 638 RSTDGVTFPQVAAGLPSSGAVGM---FHAVPGKEGDLWLAASGLYHSTNGSSWAI- 694
Db 586 VSKDTGSSF---TRG-PKLSAGTIRIDIAAHPHTAGTLYVSTDVGIIFRSTDSGTTFQGS 641

Qy 695 TCVSSAVNVGFKSAPGSSYPAVFVGTIGCVTGA---YRSDDCGTWVLINDDOHQYGNW 752
Db 642 TALNTYQIALGVGS-GSNW-NLYAFGT--GPSGARLYASGDSGASMTDIOGSGFGSID 697

Qy 753 QOAITGDHANLRRVYIGTNGRGIYV---GDIGAPSGSPSPSPSPSPSPSPSPSPSPSPSP 810
Db 698 STKVAGSGGTAGQVYVGTNGRGVFAQGTGGTGT-----SSSTKQSSSTSSA 748
```

RESULT 11

US-10-026-994-2

; Sequence 2, Application US/10026994

; Publication No. US20030113732A1





Db 338 LDRDPALDSIYLSTDAATWKDVTQLSSPSNLEGNWGHPTNAARY-KDGTVPVWLDNF 396  
Qy 413 VQP-----NPPVSP---KLGMDEAMALDPNSDRMLYGTGATLYATNDLTK-----WDS 460  
Db 397 NGPOMGGYGAPHGTPGLTKFGWMSAVLIDPFNPEHLMYGTGATLWATDLSRVEKDW-- 454  
Qy 461 GQOIHIAP---MVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSP 516  
Db 455 -----ASWYLQIDGIEENALFSLRSPKSGAALLSGIDISGMKHDLLTK-PQKMFAP 507  
Qy 517 VFTTGTSDYVLAELNPSIIRAGSPDPSSQPNDRHVAFTSDGKNW--FGSEPPGVTT-- 572  
Db 508 QFSNLDSDIDAAGNFPNVVRAGSSGHEYDSACARGAYATDGDATITPTCPGPNASHY 567  
Qy 573 -GGTVAASADGSRFVWAP--GDPGQPVVYVAVFGNSWAASQGVPA-----NAQIRSDRV 623  
Db 568 QGSTIAVDASGSIQVWSTKLDEQASGPMYSHDYGKTWS---VPAGDLKAQTANVLSDKV 623  
Qy 624 NPKTFYALNSGTFFRSTGGVTFQVAAGLPSSGAVGVMFHAPGKEDLWL-AASSGLY 682  
Db 624 QDGTFFATDGGKFFVSTDDGKSYAAKAGLVT--GTSMLPAVNPVWAGDVVVPVPEGGLF 681  
Qy 683 HSTNGGSSWSAI-TGVSAVNVGFGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDC 735  
Db 682 HSTDGASFTRVGTANATLVSAGAPKSDGKASAPSAVFIWTDKPGSDIGLYRSDN 741  
Qy 736 GTTWLINDDQHQYGNWQQAITGDHANLRRVYIGTNGRGIVYDI 780  
Db 742 GSTWRVNDQHNYSG-PTMIEADPKVYGRVYLTNGRGIVYADL 785  
RESULT 14  
US-10-395-241-14  
; Sequence 14, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent version 3.2  
; SEQ ID NO 14  
; LENGTH: 789  
; TYPE: PRT  
; ORGANISM: Geotrichum sp. M128  
US-10-395-241-14  
Query Match 21.8%; Score 1117; DB 4; Length 789;  
Best Local Similarity 35.0%; Pred. No. 2e-54;  
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;  
Qy 52 YTSNVAIGGGGFDVGVNNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDVWGNWGY 111  
Db 4 YEFKNVAIGGGGYITGIVAHPTKDLLYARTDIGAYRWDAGTSKWIPLNDFIEAQDMNI 63  
Qy 112 NGVYSIAADPNTNKVWAAVGMYT-NSWDPNDGAILRSSDQATWQITPLPKLGNWPG 170  
Db 64 MGTESIALDPNPORLYLAQQRYGVDEW-----AAFVSEDRQSFTIYESPPFMGANDMG 119  
Qy 171 RGMGERLAVDPNNNILYFGAPSGKGLWRSTDGATWQMTNFPDVGTYANPTDTTGYQ 230  
Db 120 RNNGERLAVNPFNSNEVWMTGT-EGIKSSDRAXTWTNVTISIDAF-----TNGIGYT 172  
Qy 231 SDIQGVVWVAFDKSSSSLCQASKATGIFVGVADPNPNPFWSDGGATWQAVPGAPTGF--- 287  
Db 173 S-----VIEDP-----ERNGTIYASATAPQG-MYVTHDGGVSWEPVAGQPSSWLNR 218

Qy 288 -----PHKGVFDPVNVHLXIATSNCTGSPYDGGSDYVWKFVSVTSGTWTRI- 331  
Db 219 TGAFDPKPKDAPASQPMKVALTP--NFLYVYAYDPCMGVTFGEVWRQNRRTSGAWDDIT 276  
Qy 332 -----SPVPSDTANDYFGYGLTIDRQHPNITIMVATQISWMPDTI---IFRSTDCGAT 382  
Db 277 PRVGNSSPAPYNNQTFPAGGFCGLSDVATNPRLVIT-LDRDPGALDSIYLSTDAAT 335  
Qy 383 WTRIWDWTS-----YPNRSRLRYVLDISABFWITFGVQP-----NPPVSP-- 425  
Db 336 WKDVTQLSSPSNLEGNWGHPTNAARY-KDGTVPVWLDNFNGPQWGGYGAPHGTPGLTKFG 394  
Qy 426 WNDDEAMADPFNSDRMLYGTGATLYATNDLTK-----WDSGQIHIAP---MVKGLEETA 477  
Db 395 WMSAVLIDPFNPEHLMYGTGATLWATDLSRVEKDW-----APSWYLQIDGIEENA 446  
Qy 478 VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGTSDYVLAELNPSIIVRA 537  
Db 447 ILSLSPKSGAALLSGIDISGMKHDLLTK-PQKMFAPQFSNLDSDIDAAGNFPNVVRA 505  
Qy 538 GSFPDSSQPNDRHVAFTSDGKNW--FGSEPPGVTT--GGTVAASADGSRFVWAP--G 590  
Db 506 GSSGHEYDSACARGAYATDGDATITPTCPGPNASHYQGSTIAVDASGSIQVWSTKLD 565  
Qy 591 DPGQPVVYVAVFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALNSGTFFRSTDGV 644  
Db 566 EQASGPMYSHDYGKTWS---VPAGDLKAQTANVLSDKVQDGTFFATDGGKFFVSTDDGK 621  
Qy 645 TFQVAAGLPSSGAVGVMFHAPGKEDLWL-AASSGLYHSTNGGSSWSAI-TGVSAVNV 702  
Db 622 SYAAKAGLVT--GTSMLPAVNPVWAGDVVVPVPEGGLFHTSDGASFTRVGTANATLVS 679  
Qy 703 VGFSGKS-----APGSSYPVAVFVGT--IGGVTGAYRSDCQTTWVLINDDQHQYGNWQQA 756  
Db 680 VGAPKSDGKASAPSAVFIWTDKPGSDIGLYRSDNIGSTWRVNDQHNYSG-PTMI 738  
Qy 757 TGDHANLRRVYIGTNGRGIVYDI 780  
Db 739 EADPKVYGRVYLTNGRGIVYADL 762  
RESULT 15  
US-10-395-241-18  
; Sequence 18, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent version 3.2  
; SEQ ID NO 18  
; LENGTH: 826  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Xyloglycan Oligosaccharide-Degrading Enzyme with Histidine Tag  
US-10-395-241-18  
Query Match 21.7%; Score 1114; DB 4; Length 826;  
Best Local Similarity 35.0%; Pred. No. 3.1e-54;  
Matches 281; Conservative 123; Mismatches 280; Indels 120; Gaps 31;  
Qy 52 YTSNVAIGGGGFDVGVNNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDVWGNWGY 111  
Db 5 YEFKNVAIGGGGYITGIVAHPTKDLLYARTDIGAYRWDAGTSKWIPLNDFIEAQDMNI 64



[illegible]

Search completed: March 2, 2006, 14:26:50  
Job time : 188.353 secs

November 2005

Published\_Applications Nucleic Acid and Published\_Applications Amino Acid database searches now generate two sets of results each. The Published\_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published\_Applications\_New databases; older published applications make up the Published\_Applications\_Main databases.

Searches run against Nucleic Acid Published\_Applications produce two sets of results, with the extensions **.rnpbm** (Published\_Applications\_NA\_Main) and **.rnpbn** (Published\_Applications\_NA\_New).

Searches run against Amino Acid Published\_Applications produce two sets of results, with the extensions **.rapbm** (Published\_Applications\_AA\_Main) and **.rapbn** (Published\_Applications\_AA\_New).

| Result No. | Query |       |        | DB | ID                 | Description       |
|------------|-------|-------|--------|----|--------------------|-------------------|
|            | Score | Match | Length |    |                    |                   |
| 1          | 221.5 | 4.3   | 258    | 7  | US-11-073-512-3034 | Sequence 3034, Ap |
| 2          | 219   | 4.3   | 1236   | 6  | US-10-873-528-109  | Sequence 109, App |
| 3          | 216   | 4.2   | 449    | 7  | US-11-087-099-7707 | Sequence 7707, Ap |
| 4          | 209.5 | 4.1   | 411    | 7  | US-11-036-256-89   | Sequence 89, Appl |
| 5          | 209   | 4.1   | 228    | 7  | US-11-036-256-31   | Sequence 31, Appl |
| 6          | 209   | 4.1   | 346    | 7  | US-11-036-256-33   | Sequence 33, Appl |
| 7          | 206   | 4.0   | 183    | 7  | US-11-036-256-21   | Sequence 21, Appl |
| 8          | 205   | 4.0   | 743    | 6  | US-10-485-517-351  | Sequence 351, App |
| 9          | 205   | 4.0   | 877    | 6  | US-10-485-517-200  | Sequence 200, App |
| 10         | 204.5 | 4.0   | 957    | 7  | US-11-108-172-1065 | Sequence 1065, Ap |
| 11         | 204.5 | 4.0   | 1970   | 6  | US-10-821-234-1641 | Sequence 1641, Ap |
| 12         | 198.5 | 3.9   | 534    | 6  | US-10-510-386-230  | Sequence 230, App |
| 13         | 197.5 | 3.8   | 2204   | 7  | US-11-052-554A-134 | Sequence 134, App |
| 14         | 197.5 | 3.8   | 7465   | 7  | US-11-087-099-7521 | Sequence 7521, Ap |
| 15         | 192   | 3.7   | 200    | 7  | US-11-036-256-160  | Sequence 160, App |
| 16         | 192   | 3.7   | 200    | 7  | US-11-036-256-161  | Sequence 161, App |
| 17         | 192   | 3.7   | 2000   | 7  | US-11-036-256-6    | Sequence 6, Appli |
| 18         | 192   | 3.7   | 2000   | 7  | US-11-036-256-11   | Sequence 11, Appl |
| 19         | 192   | 3.7   | 2000   | 7  | US-11-036-256-94   | Sequence 94, Appl |
| 20         | 191   | 3.7   | 231    | 7  | US-11-036-256-85   | Sequence 85, Appl |
| 21         | 190.5 | 3.7   | 397    | 7  | US-11-023-562-223  | Sequence 223, App |
| 22         | 190   | 3.7   | 252    | 7  | US-11-036-256-73   | Sequence 73, Appl |
| 23         | 189.5 | 3.7   | 499    | 6  | US-10-517-939-328  | Sequence 328, App |
| 24         | 188.5 | 3.7   | 372    | 7  | US-11-143-986-5    | Sequence 5, Appli |
| 25         | 188.5 | 3.7   | 372    | 7  | US-11-143-986-6    | Sequence 6, Appli |

|    |                       |                    |                        |                      |                    |                 |
|----|-----------------------|--------------------|------------------------|----------------------|--------------------|-----------------|
|    | Query Match           | 4.3%;              | Score 221.5;           | DB 7;                | Length 258;        |                 |
|    | Best Local Similarity | 43.6%;             | Pred. No. 7;le-06;     |                      |                    |                 |
|    | Matches               | 71;                | Conservative 15;       | Mismatches 64;       | Indels 13; Gaps 4; |                 |
| QY | 783                   | APSGGPPSVGSASPSLSP | SPSPSSSPPSPSPSSPSSSP-  | - - - - -            | PSPGFSPSPSR 836    |                 |
|    |                       |                    |                        |                      |                    |                 |
| Db | 82                    | SPSSSSSTG-S        | SPSSSSSSSSSPSSSNSSSSSS | SSSSSSSSSSSSSPSSSSSS | SSSSSS 140         |                 |
| OY | 837                   | SPSPGAASPSPSA      | SPSSSPSSSPSSPTP-       | SSSPVSGGVKVOYN-      | - - - - -          | NDSAPGDNOIK 890 |



```

; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; TITLE OF INVENTION: PEPTIDES/PROTEINS PRODUCED THEREBY
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/602,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 33
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-33

Query Match 4.1%; Score 209; DB 7; Length 346;
Best Local Similarity 43.4%; Pred. No. 4.9e-05;
Matches 49; Conservative 16; Mismatches 22; Indels 26; Gaps 6;

QY 783 APGSPSPVSP-SASPSLSPPSPSS-----SPSPSPSP-----SSPSPPSPS 825
Db 148 SPSPPTPLGLPHSPPTLSPTTPPPGPHSPPTLSPTTPPLGLPHSPPTLSPTTP 207
QY 826 PSPSPSPSPSRGSPSPSPSPS-----SSPSPPSPSPSPSPSPSPSPSPSPSG 871
Db 208 PPTPTPPGP-HAPPPPLSPSPPTPLGLPHSPPTLSPTTPP-PGVPVSG 258

RESULT 7
US-11-036-256-21
; Sequence 21, Application US/11036256
; Publication No. US20060026719A1
; GENERAL INFORMATION:
; APPLICANT: XU, JIANFENG
; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND
; TITLE OF INVENTION: PEPTIDES/PROTEINS PRODUCED THEREBY
; FILE REFERENCE: 27211/04130
; CURRENT APPLICATION NUMBER: US/11/036,256
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: 60/602,562
; PRIOR FILING DATE: 2004-08-18
; PRIOR APPLICATION NUMBER: 60/582,027
; PRIOR FILING DATE: 2004-06-22
; PRIOR APPLICATION NUMBER: 60/536,486
; PRIOR FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 21
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid construct
US-11-036-256-21

Query Match 4.0%; Score 206; DB 7; Length 183;
Best Local Similarity 44.2%; Pred. No. 3.5e-05;
Matches 50; Conservative 16; Mismatches 23; Indels 24; Gaps 7;

QY 783 APGSPSPVSP-SASPSLSPPSPSS-----SPSPSPSP-----SSPSPPSPS 825
Db 72 SPSPPTPLGLPHSPPTLSPTTPPPGPHSPPTLSPTTPP-PGVPVSG 131

```

**Qy** 826 pspapspspgrspspaspspsppss--spssspss-spsspsspspspsvsggvkv 875  
|:| | | | | | | | : | | | | | | | |  
**Db** 132 ftfptpppgp-hsfpplsluspfstfplglghspfppluspfpntppp---gvpv 180

RESULT 8  
US-10-485-517-351  
; Sequence 351, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:

```

; ; ; PRIOR FILING DATE: 2002-01-09
; ; ; NUMBER OF SEQ ID NOS: 424
; ; ; SOFTWARE: PatentIn version 3.1
; ; ; SEQ ID NO 200
; ; ; LENGTH: 877
; ; ; TYPE: PRF
; ; ; ORGANISM: Staphylococcus aureus
US-10-485-517-200

```

Query Match 4.0%; Score 205; DB 6; Length 877;  
Best Local Similarity 23.4%; Pred. No. 0.00024;  
Matches 108; Conservative 43; Mismatches 179; Indels 132; Gaps 19

Qy 507 AVPSITFTSPVFTT---GTSVDYAEINPSIIVRAGSFDPSQPNDRHVAFSTDGK--- 559

Db 194 AVPSLVAEPVNAADAAGTINVNDKVTASNFLEKTTFDPNQSGNTFWAAFTVTDKYKS 253

QY 560 -NWFGSEPGVTTGGTVAASADGSREFWAP-----GDPGQPVVYV-----GFG 603  
 Db 254 GDYFTAKPLDSTLGTGNDVYDSNSNNTMPIADIKSTNGDVAKATYDILTKTYTFVFTDYV 313

Qy 604 NSWAASQG-----VPANAQRSDRVNPKTFVSLNGTYYRSTGCVT 645  
| : | | | | | | | | | | | | | | | |  
Db 314 NKNENGGFSLPFTTRAKAPKSGTYDANIADENFNKKITYN-----YSSPIAGID 367  
| : | | | | | | | | | | | | | | | |

[illegible]

RESULT 8  
US-10-485-517-351  
; Sequence 351, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:

```

; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02

```

```

; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09

```

```

; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 351
; LENGTH: 743
; TYPE: prt

```

US-10-485-517-351 ORGANISM: Staphylococcus aureus  
Query Match 4.0% Score 205 DB 6 Length 743

Best Local Similarity 23.4%; Pred. No. 0.00019;  
Matches 108; Conservative 43; Mismatches 179; Indels 132; Gaps 19;  
QY 507 AVPTSTLTFTSPVFTT----GTSVDYAEINPSIIIVRAGSFDPSQNDRRHVAFTDGGK--- 559







QY 765 -----RVYIGN-GRGIV-----YDGIIGAPSGSPSPSVSPSASPSLSFSPSPSS 808  
 :  
 Db 786 PEALTEGVDPIDPIHIPINIDAGVVTLCQFSIVAENNIDFTIIPITINITLPTAAITVG 845  
 :  
 QY 809 SPSPSPSPSPSPSPSPSP-SFSPSPSRPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSS 867  
 :  
 Db 846 GPTISIGITASIGISITIBIDIIPATSGRGNSTTSPSGFFNSGAGSAGFLNVAVGAS 905  
 :  
 QY 868 PVSGGVKVKYNKNDSPEDNQIKPLGOVNNTG-----SSVDLSSTTVRYWFTRDGG 919  
 :  
 Db 906 GISG-----YLVNVAL-----GSGVTNVGHVSFYNASALDLVTPAFASGLARD-G 951  
 :  
 QY 920 SSTLVVNCDDWAICCGNIRASFGS 943  
 :  
 Db 952 MGTMTNLNLGLANLGNN--AGFGN 973  
 :

RESULT 14  
 US-11-087-099-7521  
 ; Sequence 7521, Application US/11087099  
 ; Publication No. US20060041961A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S. et al.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53450)B EP  
 ; CURRENT APPLICATION NUMBER: US/11/087,099  
 ; CURRENT FILING DATE: 2005-03-22  
 ; NUMBER OF SEQ ID NOS: 12464  
 ; SEQ ID NO 7521  
 ; LENGTH: 7465  
 ; TYPE: PRT  
 ; ORGANISM: Magnetococcus sp. MC-1  
 US-11-087-099-7521

Query Match 3.8%; Score 197.5; DB 7; Length 7465;  
 Best Local Similarity 18.8%; Pred. No. 0.007;  
 Matches 211; Conservative 105; Mismatches 461; Indels 346; Gaps 41;

QY 17 VSLIAATAS-----FAVAALGLVLPIAITASPAAHAATTQPYTWSNVAIGGG----- 63  
 :  
 Db 6460 VSLTSGSSASLNEDDTTSVSGTLTVVDADVIDATVTAATSVG-TYGTFSVNGSVMSYQ 6518  
 :  
 QY 64 -----FVDGIVFNEGAPGILYVRTDIGMY-----RWDAANGRWIPLLDWVGWNW 109  
 :  
 Db 6519 LDNSKAVVOGLTQGQQVSESFVHTSDGGSQSILTMTITGRQDSA----- 6562  
 :  
 QY 110 GYGCVVSIADPINTNKVWAAVMGYT-NSWDPNDAAILRSSDOGA-----TWQ 156  
 :  
 Db 6563 ----VWGSCTGVTEDSTLSSGMLSVSDADSGEAAVAVASTQTGYSFSLDSGWSAQY 6618  
 :  
 QY 157 ITPLPKLGCNMPCRGMERL---AYDPNNDNIL----- 187  
 :  
 Db 6619 LNNASQVQALIMGQQVSESFVATVDGTASSVVTVWGQAQAQLGGVQSGGVVEDGQL 6678  
 :  
 QY 188 -----YFCAPSCKGLWR-----STDGATWSOMTFPDVPDVGTIANPTDTTGY 229  
 :  
 Db 6679 STAGLTISADAGOSLYQVASTDMGYHPAMDSSGNWS-----YVLN-NELAAI 6727  
 :  
 QY 230 QSDIQG-VVWAFDKSSSLGQASKITFVGAVDPNNPVFWSRGG---ATWQAVPGAPT 285  
 :  
 Db 6728 QOMIAGQVLESFETSSSDGTQAVSVTI-TGSQDNNAVITGEAAGQVSETNDDSQSMATG 6786  
 :  
 QY 286 FIPKHGVDFPVHVLVIATSNTPGDGSGDWKESVT-SGTWTRI-----SPVPS 336  
 :  
 Db 6787 KLINISDLDSQQAHFV-----GGPOAQOYQ---AFTLQDGDNWTVYLDNHGVAMAGLGT 6836  
 :  
 QY 337 TDANDYFGYSGLTIDROHNPTIMVATQISWPDTIIFRSTDCGATWTRIDWTSYPNRS 396  
 :  
 Db 6837 GDVVTEFTVRSI--DNAETTV-----TTINGLEDPPPAEIEEARLAEOR 6883  
 :  
 QY 397 LRYVLD-----ISAEPWLTFGPVQNPVPSPKLGWMDEMAIDPPFNDRMLYGTGA 447  
 :

RESULT 15  
 US-11-036-256-160  
 ; Sequence 160, Application US/11036256  
 ; Publication No. US20060026719A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KIELISZEWSKI, MARCIA  
 ; APPLICANT: XU, JIANPENG  
 ; TITLE OF INVENTION: METHODS OF PRODUCING PEPTIDES/PROTEINS IN PLANTS AND  
 ; TITLE OF INVENTION: PEPTIDES/PROTEINS PRODUCED THEREBY  
 ; FILE REFERENCE: 27211/04130  
 ; CURRENT APPLICATION NUMBER: US/11/036,256  
 ; CURRENT FILING DATE: 2005-01-14  
 ; PRIOR APPLICATION NUMBER: 60/602,562  
 ; PRIOR FILING DATE: 2004-08-18  
 ; PRIOR APPLICATION NUMBER: 60/582,027  
 ; PRIOR FILING DATE: 2004-06-22  
 ; PRIOR APPLICATION NUMBER: 60/536,486  
 ; PRIOR FILING DATE: 2004-01-14  
 ; NUMBER OF SEQ ID NOS: 173  
 ; SOFTWARE: PatentIn Ver. 3.3  
 ; SEQ ID NO 160  
 ; LENGTH: 200  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 ; OTHER INFORMATION: amino acid motif  
 ; FEATURE:  
 ; NAME/KEY: MOD\_RES  
 ; LOCATION: (1)..(1)

|    |                    |                                 |
|----|--------------------|---------------------------------|
| 1  | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 2  | FEATURE:           |                                 |
| 3  | NAME/KEY:          | MOD RES                         |
| 4  | LOCATION:          | (3) .. (3)                      |
| 5  | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 6  | FEATURE:           |                                 |
| 7  | NAME/KEY:          | MOD RES                         |
| 8  | LOCATION:          | (5) .. (5)                      |
| 9  | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 10 | FEATURE:           |                                 |
| 11 | NAME/KEY:          | MOD RES                         |
| 12 | LOCATION:          | (7) .. (7)                      |
| 13 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 14 | FEATURE:           |                                 |
| 15 | NAME/KEY:          | MOD RES                         |
| 16 | LOCATION:          | (9) .. (9)                      |
| 17 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 18 | FEATURE:           |                                 |
| 19 | NAME/KEY:          | MOD RES                         |
| 20 | LOCATION:          | (11) .. (11)                    |
| 21 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 22 | FEATURE:           |                                 |
| 23 | NAME/KEY:          | MOD RES                         |
| 24 | LOCATION:          | (13) .. (13)                    |
| 25 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 26 | FEATURE:           |                                 |
| 27 | NAME/KEY:          | MOD RES                         |
| 28 | LOCATION:          | (15) .. (15)                    |
| 29 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 30 | FEATURE:           |                                 |
| 31 | NAME/KEY:          | MOD RES                         |
| 32 | LOCATION:          | (17) .. (17)                    |
| 33 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 34 | FEATURE:           |                                 |
| 35 | NAME/KEY:          | MOD RES                         |
| 36 | LOCATION:          | (19) .. (19)                    |
| 37 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 38 | FEATURE:           |                                 |
| 39 | NAME/KEY:          | MOD RES                         |
| 40 | LOCATION:          | (21) .. (21)                    |
| 41 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 42 | FEATURE:           |                                 |
| 43 | NAME/KEY:          | MOD RES                         |
| 44 | LOCATION:          | (23) .. (23)                    |
| 45 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 46 | FEATURE:           |                                 |
| 47 | NAME/KEY:          | MOD RES                         |
| 48 | LOCATION:          | (25) .. (25)                    |
| 49 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 50 | FEATURE:           |                                 |
| 51 | NAME/KEY:          | MOD RES                         |
| 52 | LOCATION:          | (27) .. (27)                    |
| 53 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 54 | FEATURE:           |                                 |
| 55 | NAME/KEY:          | MOD RES                         |
| 56 | LOCATION:          | (29) .. (29)                    |
| 57 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 58 | FEATURE:           |                                 |
| 59 | NAME/KEY:          | MOD RES                         |
| 60 | LOCATION:          | (31) .. (31)                    |
| 61 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 62 | FEATURE:           |                                 |
| 63 | NAME/KEY:          | MOD RES                         |
| 64 | LOCATION:          | (33) .. (33)                    |
| 65 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 66 | FEATURE:           |                                 |
| 67 | NAME/KEY:          | MOD RES                         |
| 68 | LOCATION:          | (35) .. (35)                    |
| 69 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |
| 70 | FEATURE:           |                                 |
| 71 | NAME/KEY:          | MOD RES                         |
| 72 | LOCATION:          | (37) .. (37)                    |
| 73 | OTHER INFORMATION: | Lys, Ser, Ala, Thr, Gly, or Val |

|  |                   |  |
|--|-------------------|--|
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (39)..(39)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (41)..(41)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (43)..(43)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (45)..(45)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (47)..(47)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (49)..(49)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (51)..(51)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (53)..(53)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (55)..(55)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (57)..(57)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (59)..(59)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (61)..(61)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (63)..(63)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (65)..(65)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (67)..(67)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (69)..(69)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (71)..(71)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   | NAME/KEY: MOD RES |  |
| LOCATION: (73)..(73)                               |                   |  |
| OTHER INFORMATION: Lys, Ser, Ala, Thr, Gly, or Val |                   |  |
| FEATURE:   |                   |  |



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:06:58 ; Search time 345.215 Seconds  
(without alignments)  
4927.737 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5134

Sequence: 1 MDRSENRLTWRRRLVSL.....RASFGSVNPATPTADTYLQX 957

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q-/abss/ABSSWEB\_spool/US09917376/runat 02032006\_091454\_8101/app\_query.fasta\_1  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abss06h -USER=US09917376 @CGN 1.1 278 @runat 02032006\_091454\_8101  
-NCPU=6 -ICPU=3 -NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/1/ina/1 COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCUS COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq.\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq.\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID               | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 409   | 8.0         | 11707  | US-09-136-574A-1    | Sequence 1, Appli |
| 2          | 392.5 | 7.6         | 1103   | US-09-533-559-7511  | Sequence 7511, Ap |
| C 3        | 358   | 7.0         | 19383  | US-09-949-016-16031 | Sequence 16031, A |
| C 4        | 345   | 6.7         | 18596  | US-09-318-448-11    | Sequence 11, Appl |
| C 5        | 345   | 6.7         | 18596  | US-09-577-286-11    | Sequence 11, Appl |
| C 6        | 345   | 6.7         | 18597  | US-09-962-665-8     | Sequence 8, Appli |
| C 7        | 345   | 6.7         | 18597  | US-09-963-333-8     | Sequence 8, Appli |
| C 8        | 345   | 6.7         | 18597  | US-09-962-677-8     | Sequence 8, Appli |
| C 9        | 334   | 6.5         | 6416   | US-09-136-574A-2    | Sequence 2, Appli |

|      |       |     |         |   |                      |                   |
|------|-------|-----|---------|---|----------------------|-------------------|
| 10   | 319.5 | 6.2 | 8211    | 3 | US-09-252-991A-13656 | Sequence 13656, A |
| 11   | 296.5 | 5.8 | 4767    | 3 | US-09-410-551B-28    | Sequence 28, Appl |
| 12   | 296.5 | 5.8 | 4767    | 3 | US-09-940-316B-28    | Sequence 28, Appl |
| 13   | 293   | 5.7 | 4818    | 3 | US-09-410-551B-32    | Sequence 32, Appl |
| 14   | 293   | 5.7 | 4818    | 3 | US-09-940-316B-32    | Sequence 32, Appl |
| C 15 | 288.5 | 5.6 | 4131    | 3 | US-09-252-991A-13773 | Sequence 13773, A |
| C 16 | 288.5 | 5.6 | 4571    | 3 | US-09-410-551B-18    | Sequence 18, Appl |
| 17   | 288.5 | 5.6 | 4571    | 3 | US-09-940-316B-18    | Sequence 18, Appl |
| 18   | 288   | 5.6 | 288     | 3 | US-09-119-507B-111   | Sequence 111, App |
| 19   | 288   | 5.6 | 288     | 3 | US-09-547-693-111    | Sequence 111, App |
| C 20 | 282.5 | 5.5 | 441529  | 3 | US-09-103-840A-1     | Sequence 1, Appli |
| C 21 | 281.5 | 5.5 | 4403765 | 3 | US-09-103-840A-2     | Sequence 2, Appli |
| C 22 | 280   | 5.5 | 3129    | 3 | US-09-252-991A-13873 | Sequence 13873, A |
| 23   | 279   | 5.4 | 4547    | 3 | US-09-410-551B-22    | Sequence 22, Appl |
| 24   | 279   | 5.4 | 4547    | 3 | US-09-940-316B-22    | Sequence 22, Appl |
| 25   | 273   | 5.3 | 4674    | 3 | US-09-410-551B-26    | Sequence 26, Appl |
| 26   | 273   | 5.3 | 4674    | 3 | US-09-940-316B-26    | Sequence 26, Appl |
| 27   | 269   | 5.2 | 4478    | 3 | US-09-410-551B-16    | Sequence 16, Appl |
| 28   | 269   | 5.2 | 4478    | 3 | US-09-940-316B-16    | Sequence 16, Appl |
| C 29 | 268.5 | 5.2 | 31467   | 3 | US-09-949-016-13134  | Sequence 13134, A |
| C 30 | 268.5 | 5.2 | 31868   | 3 | US-09-949-016-11907  | Sequence 11907, A |
| C 31 | 268   | 5.2 | 31391   | 3 | US-09-949-016-14319  | Sequence 14319, A |
| C 32 | 266   | 5.2 | 4188    | 3 | US-09-252-991A-13774 | Sequence 13774, A |
| 33   | 266   | 5.2 | 77536   | 3 | US-09-410-551B-1     | Sequence 1, Appli |
| 34   | 266   | 5.2 | 77536   | 3 | US-09-940-316B-1     | Sequence 1, Appli |
| C 35 | 265.5 | 5.2 | 77536   | 3 | US-09-410-551B-1     | Sequence 1, Appli |
| C 36 | 265.5 | 5.2 | 77536   | 3 | US-09-940-316B-1     | Sequence 1, Appli |
| 37   | 265   | 5.2 | 4737    | 3 | US-09-410-551B-30    | Sequence 30, Appl |
| 38   | 265   | 5.2 | 4737    | 3 | US-09-940-316B-30    | Sequence 30, Appl |
| 39   | 264.5 | 5.2 | 6210    | 3 | US-10-212-962-1      | Sequence 1, Appli |
| 40   | 263   | 5.1 | 150394  | 3 | US-09-914-286-1      | Sequence 1, Appli |
| 41   | 262   | 5.1 | 150394  | 3 | US-09-949-016-13042  | Sequence 13042, A |
| C 42 | 260   | 5.1 | 3337    | 2 | US-08-072-610-1      | Sequence 1, Appli |
| C 43 | 260   | 5.1 | 3337    | 2 | US-08-719-822B-1     | Sequence 1, Appli |
| C 44 | 260   | 5.1 | 3337    | 3 | US-09-092-458-1      | Sequence 1, Appli |
| C 45 | 260   | 5.1 | 3337    | 3 | US-08-719-821C-1     | Sequence 1, Appli |

#### ALIGNMENTS

RESULT 1

US-09-136-574A-1

; Sequence 1, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing  
; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

Fabrics Using Truncated

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 11707 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-136-574A-1

Alignment Scores:

Pred. No.: 5,45e-11 Length: 11707  
Score: 409.00 Matches: 263  
Percent Similarity: 32.4% Conservative: 163  
Best Local Similarity: 20.0% Mismatches: 386  
Query Match: 8.0% Indels: 504  
DB: 3 Gaps: 58

US-09-917-376-1 (1-957) x US-09-136-574A-1 (1-11707)

```

QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
   ::::::::::::::::::::
DB 735 AAGAGGGTAATTCATCTT-----TCCTTATGTTTTTTTAAACACGCTGTAGGT 791
QY 34 ValLeuProIleAlaThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr 53
   ::::::::::::::::::::
DB 792 ACTTTG-----ATATTCATCAGGAGCAAAACAGCAGCA-----TATACT 833
QY 54 TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly 73
   ::::::::::::::::::::
DB 834 -----GTTGATTTTGAAGGT 848
QY 74 AlaProGlyIleLeuTyr-----ValArgThrAspIleGlyGly 86
   ::::::::::::::::::::
DB 849 GCTGATACTTTATCTTACTTCTTATGTAAGAAATCGAGCATAGCAGTTGACATGGGCAAT 908
QY 87 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 106
   ::::::::::::::::::::
DB 908 ----- 908
QY 107 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 126
   ::::::::::::::::::::
DB 909 -----GCATATAATGCTAAAGTAGTGTCTCAGGGTGCA-----AATAGAAGTTCA 953
QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 146
   ::::::::::::::::::::
DB 954 ATATGGGATGGAGTTGCAGTT-----GAGGTTTAAA 983
QY 147 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThr----- 158
   ::::::::::::::::::::
DB 984 AACATTATGACAAATGAACACATGGTAGTTTCACGGTAGTAAACATAGTACCAG 1043
QY 159 ---ProLeuProPheLysLeu-----GlyGlyAsnMetProGlyArg 171
   ::::::::::::::::::::
DB 1044 AACCCGGTTGCAATTTGGTATCTCAGCGGTTTACGACGATCGAAGTGGGTAAAGACTACT 1103
QY 172 GlyMetGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 191
   ::::::::::::::::::::
DB 1104 CTATAGGTGAGGTTGTGGCTATT-----CCAAAT----- 1133
QY 192 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 211
   ::::::::::::::::::::
DB 1134 -----TATTGGAAGAAAATT--- 1148
QY 212 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 231

```

```

DB 1149 -----GTTGGTAAATGGACTCCAAATATTAGCAATGTCTCAGGAATTTGTTAATT 1196
QY 232 AspileGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 251
   ::::::::::::::::::::
DB 1197 GTAATACACACAATTTGTAGAAGCAAGTAGATTATTAATGTTGAC----- 1241
QY 252 SerLysThrIlePheValGlyValAlaAspProAsn-----AsnProValPhe 267
   ::::::::::::::::::::
DB 1242 -----TATATCCAAATATGATGATGATATGTTACCTATCAAAATGCAAGTGACA 1289
QY 268 TrpSerArg-----AspGlyGlyAlaThr-----TrpGlnAla-----Val 279
   ::::::::::::::::::::
DB 1290 TTTTCAAGTGGATTGAAAGTGGCACTACCGAGGTTGGCAGGCAAGGGAGCGGTGT 1349
QY 280 ProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisVal 299
   ::::::::::::::::::::
DB 1350 ACAGTAAACACAGTAGCGTTGTGCATATATAGTGCAGATATAGT----- 1394
QY 300 LeuTyrIleAla-----ThrSerAsnThrGlyGly-----ProTyrAspGly----- 313
   ::::::::::::::::::::
DB 1395 TTGTACGTCACTGGGAAGAACGTCAAATTTGGCATGTGTGCACAGATTCGGGTAGATACAATT 1454
QY 314 ---SerSerGlyAspValTrpLysPheSerVal-----ThrSerGlyThrTrp 328
   ::::::::::::::::::::
DB 1455 TTGGAACAGGGTAAAGTGTATAAAATAAGTGTGTTGGTTTATCAGAACAGTGTGTTCAACT 1514
QY 329 ThrArgIleSerProValProSerThrAspThrAlaAsnAsp-----TyrPheGlyTyrSer 347
   ::::::::::::::::::::
DB 1515 CAAAATATGTCATTAACTATCACAAGAAAGATTGTGTCACAGACTCTTCAACAAGCTATGAA 1574
QY 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367
   ::::::::::::::::::::
DB 1575 AATCTGATATATACAGGGATGTACCGAGTAATACTCGTGGTTGAGCTGAGTGAAGCTAC 1634
QY 368 TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387
   ::::::::::::::::::::
DB 1635 -----TCAATTCCTGCTGGTGTACAGTTAGCGAGTGTGTTG 1670
QY 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyr-----ValLeu 401
   ::::::::::::::::::::
DB 1671 CTTTATGTGAGGCACAAAATGCAAAATTTGGCTTTCTCGGTTGATGATTTAAAGATTAT 1730
QY 402 AspileSer-----AlaGluPro----- 407
   ::::::::::::::::::::
DB 1731 GATTATCCAAGTTGGCTGAACCTGAATGGGAGATACCATCTTTGATAGAAAGATATAGA 1790
QY 407 ----- 407
DB 1791 GATTATTTCAAAGTAGGAGTAGCTTTGTCTTACAAAAGCATTGCTCTGATACAGAAAAG 1850
QY 407 ----- 407
DB 1851 AAGATGGTTTTGAAGCATTTTCAATAGTATTACTGCGGGAACGAAATGAACCATCAGAG 1910
QY 408 -----TrpLeuThrPheGly 412
   ::::::::::::::::::::
DB 1911 TTACTTGTGATGAAAATACTTACAACTTTAGCAAGCAGACGAAATTTGTAATTTTGA 1970
QY 413 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAla----- 430
   ::::::::::::::::::::
DB 1971 ACAAGTAACAACATGGCCATCAGAGGTACATACACTGTTTGGCATGAGCAAAACCCGAC 2030
QY 430 ----- 430
DB 2031 TGGTTTTTCAAGGACACAAATGGAATACGTTTGACGAAGGATGCATTTGTAAGCAGATTA 2090
QY 430 ----- 430
DB 2091 AAACAGTATATTTATACGGTAGTGGGAAGATATAAAGGAAGGTTTATGATGGGATGTG 2150
QY 431 -----MetAlaIleAspProPheAsnSerAspArgMetLeu----- 442
   ::::::::::::::::::::

```

Db 2151 GTAAATGAAGCAATAGATGAAGAGTCAAGGTGATGATTCAGGAGATCTAACTGGTACAAC 2210  
Qy 443 -----TyrGlyThrClyAlaThrLeuTyrAla----- 451  
Db 2211 ATTTGTAGTCCCGAATATATTGAGAAGCTTTTATATGGGCACATGAAGCCGATCCAGAC 2270  
Qy 452 -----ThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleAla 467  
Db 2271 GCAAAATTTGTTTACACAGATTACACACAGAAACAGTCAAGAGACAGTTTATTATTC 2330  
Qy 468 ProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly 487  
Db 2331 AACATGATTAGAGTCTCAAGGAAAAGGTGT-----CCATTTCATGGA 2375  
Qy 488 AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAla 507  
Db 2376 ATAGGATGCGAGAGTCAATA-----AATCTTGATTGGCCCTCGATTAGCGAG 2423  
Qy 508 ValProSerThrIle-----PheThrSerProValPheThrGlyThrSerValAsp 525  
Db 2424 ATAGAGAACACCATTAAGATTCTTCAGCTCT-----ATACCTGGATTGGAGATACAC 2474  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545  
Db 2475 ATTACGGAGCTTGATAGATGATTTTATCAGTGGGTTCCGAGTACCAGTTACTCAACGCCA 2534  
Qy 546 ProAsnAsp----- 548  
Db 2535 CCAAGAGATCTCCTCATAAACAGCAATGAGATATAGGAGTTATTGATTGTTGTTAA 2594  
Qy 549 -----ArgHisValAlaPhe----- 553  
Db 2595 AAGTACAACAATGTAATAACAAGTGAACATCTGGGAGTGAAGGATGATTACTCATGG 2654  
Qy 554 ---SerThrAspGlyGlyAsn----- 560  
Db 2655 CTGAGTCAAAATTTGGAAAAGTGATTACCGTGTGTTATTGTTGATAAACTATAAATCA 2714  
Qy 561 -----TrpPheGlnGlySerGluProGlyGlyValThrGlyThrGlyThrValAla 577  
Db 2715 AAATATGCTTTTGGAGCTGATTGAGCCAACTGTGATACCGGCCAACTCAACATTGCCA 2774  
Qy 578 Ala-----SerAlaAspGlySerArgPheVal 586  
Db 2775 GCACCACCACTATTCAAAATACCTACACCACTCCACACCAACCCGACACGACATG 2834  
Qy 587 TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp 606  
Db 2835 AGTGCACGCCAACACACGACA---CCGACGGCATCACCGGTAGGT---GGCAGTTTACTGG 2888  
Qy 607 AlaAlaSerGlnGlyValProAla-----AsnAlaGlnIle 618  
Db 2889 ACGCGAGTAGAGATTACAGTGCCTGAAGGTATGGTATGGAATGGGAATTTAAGCAGC 2948  
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsn---GlyThr----- 635  
Db 2949 CCGACGAATGATTGAATCCTAAG-----ATAAGATAGAGATGTTGGAGACGACAGC 3002  
Qy 636 -----PheTyrArgSerThrAspGlyGlyValThr--- 645  
Db 3003 GTAGATCTTAGCAGGGTGAAGGTAAAGTACTGTTGATACAGATAGATGGTGAGCAACACAG 3062  
Qy 646 -----PheGlnPro-----ValAlaAla 651  
Db 3063 AGTGAAGTGTAAACAGCAGCATAGATCCTCGGTATATAGATGTGAAGTTTGTGAAGCTT 3122  
Qy 652 GlyLeuProSerSerGlyVala-----ValGly 660  
Db 3123 GGAGCGAAGCAGCGCGGACCGGATTACTATGTGGAGATAGGCTTTAAGAGTGGAGCAGCG 3182  
Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
Db 3183 GTTTTGGCAGCAGGGCAAGAACACGAGGAGATAGACTTAGCATACAGAAAGGCGAGTGGC 3242

Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla----- 693  
Db 3243 AGCTACAATCACTCAAAATCACTATTTCGGTGAAGAGTGCACAGGCTATATAGAGAACGAG 3302  
Qy 694 ---IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySer 712  
Db 3303 AAGGTAAACAGGTATATAGATGATGATCTTGTATGGGGAAGAGCGCGAGCAGAACGCC 3362  
Qy 713 SerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla----- 729  
Db 3363 CAGATCAAGGTATGTTATCGGAATGGAATTTAAGCAGCGCCGACGAATGTTATTGAATCCT 3422  
Qy 730 ---TyrArgSerAspAspCysGlyThrThr----- 738  
Db 3423 AAGATAAAGATAGAGAATTTGCGACGACACGCGGTAGATCTTTAGCAGGTTGAAGGTAAGA 3482  
Qy 739 ---TrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThr 757  
Db 3483 TACTGGTACACATAGATGGTGAGGCAACACAGAGTGTAAAGTGTAAACAGCAGCATAAAC 3542  
Qy 758 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIle---Val 776  
Db 3543 CCTGCGTATATAGATGTGAAGTTTGTGAAGCTTGGAGCAANTGCCAGTGGAGCGGATTAC 3602  
Qy 777 TyrGlyAspIleGly-----GlyAla----- 783  
Db 3603 TATGTGGAGATAGGCTTTTAAGAGTGGAGCAGGGTTTTGGCAGCAGGGCAGCAGCAG 3662  
Qy 784 -----ProSerGlySer----- 787  
Db 3663 GAGATAAGACTTTAGCATACAGAAGGCGCAGTGGCAGCTACAANTCAGTCAAAATGACTATTTCG 3722  
Qy 787 ----- 787  
Db 3723 GTGAGGAGTGCACAGGCTATATAGAGAAGAGTAAACGGGTATATAGATGGTGGC 3782  
Qy 788 -----ProSerProSerValSerProSerAlaSerProSerLeuSer 801  
Db 3783 ATAGTGTGGGGAAGAGAGCGGAGGTACAAGCCGCGGGAGTAGTAGAACCCGACA 3842  
Qy 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 821  
Db 3843 CCGGCACCGACCCCGACATCGACGCGCACCAACACCTACAACACCACTGCACCGACA 3902  
Qy 822 SerSerProSerProSerProSerProSerProSerProSerProSerProSer 841  
Db 3903 TCAGCCCGACACCGAGCCCAACAGTGCAGCAACCGCCGACTCCAACGCGCAGCCGACA 3962  
Qy 842 AlaSerProSerProSerProSerProSerProSerProSerProSerProSer 861  
Db 3963 GTGAGGTTACTGTGACTCCGACCCGACCA-----ACACCGACG 4004  
Qy 862 ProThrProSerSerProValSerGly---GlyValLysValGlnTyrLysAsnAsn 880  
Db 4005 CCGACACCGACGAGGACACCTGCGCAGGAGTGGTTGAGGTACTATACAGAACAAAT 4064  
Qy 881 AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900  
Db 4065 GAGACAAGTGGCAGCACAAAGTCTATAGGCGGTGTTTAAAGATAGTAGTAATGGAGGCGAGC 4124  
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920  
Db 4125 AGCAGTGTGTATCTTAGCAGGGTTAAGATAAGTACTGTGTACACATGATGGTGTGACAAG 4184  
Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940  
Db 4185 CCACAGAGTGGCGTA---TGTCACCTGGGCACAGATAGGGGCAAGCAATGTGACATTCAAT 4241  
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 4242 TTTTGTGAAGCTGAGCAGCGGAGTGTGAGTGGCGGGATTATTACTTTGGAG 4289

## RESULT 2

US-09-533-559-7511  
 ; Sequence 7511, Application US/09533559  
 ; Patent No. 6902887  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Randy M. Berka  
 ; APPLICANT: Michael W. Rey  
 ; APPLICANT: Jeffrey R. Shuster  
 ; APPLICANT: Sakari Kauppinen  
 ; APPLICANT: Ib Groth Clausen  
 ; APPLICANT: Peter Bjarke Olsen  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 5849,200-US  
 ; CURRENT APPLICATION NUMBER: US/09/533,559  
 ; CURRENT FILING DATE: 2000-03-22  
 ; EARLIER APPLICATION NUMBER: 09/273,623  
 ; EARLIER FILING DATE: 1999-03-22  
 ; NUMBER OF SEQ ID NOS: 7860  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7511  
 ; LENGTH: 1103  
 ; TYPE: DNA  
 ; ORGANISM: Tricoderma reesei  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(1103)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-533-559-7511

## Alignment Scores:

Pred. No.: 2,44e-11 Length: 1103  
 Score: 392.50 Matches: 114  
 Percent Similarity: 45.0% Conservative: 36  
 Best Local Similarity: 34.2% Mismatches: 125  
 Query Match: 7.6% Indels: 58  
 DB: 3 Gaps: 10

US-09-917-376-1 (1-957) x US-09-533-559-7511 (1-1103)

QY 348 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 367  
 Db 3 GCGCTTGGCGCTCGAATTTGCAAAAGCCAGGAAACCTTTGTTGCTTCTTGAACCTCTTGG 62  
 QY 368 TrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 387  
 Db 63 TGCCAGAGTCTCAGCTGTTTCGTCGACCGACTCTGGGACACATGGAGCCCGAATCTGG 122  
 QY 388 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 407  
 Db 123 GCGTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAGACCG 182  
 QY 408 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 422  
 Db 183 TGGATCAAGAACAACATTATCGATGTGACGAGCGAGTCACCGCTCCGATGCTCTNATCAAG 242  
 QY 423 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 442  
 Db 243 CGCCTCGGCTGGATGATTGAGTCTNTCGAGATTGACCCACCCACCAANNACTGCGCTT 302  
 QY 443 -TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer----- 460  
 Db 303 TTACGGCACCGGATGACATNTTTGGCGCCACGATTTCCCACTGGGACACGCGCCC 362  
 QY 461 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluGluThrAlav 478  
 Db 363 ACAATGTGTCAATCAATTAATCTTGGCAGACGGGATTCGAAGGAATT-----TTTCGGT 416  
 QY 478 alaAsnAspLeuIleSerProProSerGly----- 487  
 Db 417 TCAAGGACCTGGCGCTTTTCCACCGGGGGGAAGCGAGCTTTTGGCCGCAAGTCCGGAGACG 476  
 QY 488 -----AlaProLeuIleSerAlaLeuAspLeuGlyGlyPheThrHisAlaAspV 505

Db 477 ANCAACGGGTTTACCTTTTGCCGACAGAAACGACCTTGGG----- 517  
 QY 505 alThrAlaValProSerThrIlePheThrSerProValPheThrThrThrGlyThr-SerVal 524  
 Db 518 -----ACATTGCCGACGAGCGTTTGGGCAACTCCCATCATGGCCACCTCGACGAGCGTC 572  
 QY 525 AspTyrAla--GlueuAsnProSerIleIleValArgAlaGlySer----- 539  
 Db 573 GACTAGCGCGGAACTCGGTCAAGAGCGCTTCTCGGTCGGCAACACCGCGCGGCACG 632  
 QY 540 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 554  
 Db 633 CAACAAGGTGGCCATTATTTTCCGAACGGCGCGCGCATCTCTGGTCACCGCTCGTCCGGC- 686  
 QY 554 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG 574  
 Db 687 GAACTAACGCT-----GGNTCCGAACACCGCTTTTCCATTGAAACGGCG 728  
 QY 574 lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 594  
 Db 729 GCGCGTGGCTATTTCGCGCGACGCGACGATCTCTGGTCACCGCTCGTCCGGC- 787  
 QY 594 lnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 614  
 Db 788 -----GTGACGCGCTCGCAGTTTCCAGGGCAGCTTTGCTCTCGAGCCTCGCCG 839  
 QY 614 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG 634  
 Db 840 CGGGCGCGCTCATCGNCTCGGACAAAGAGACCAACAGCGTNTTCTACGCGCGCTCGGAT 899  
 QY 634 lyThrPheTyrArgSerThrAspGlyGly 643  
 Db 900 CGACCTTTTACGTACGACGACACCGCGC 928

## RESULT 3

US-09-949-016-16031/c  
 ; Sequence 16031, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CL001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; CURRENT FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16031  
 ; LENGTH: 19383  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)...(19383)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-16031

## Alignment Scores:

Pred. No.: 3.36e-08 Length: 19383  
 Score: 358.00 Matches: 71  
 Percent Similarity: 84.9% Conservative: 2  
 Best Local Similarity: 82.6% Mismatches: 13  
 Query Match: 7.0% Indels: 0  
 DB: 3 Gaps: 0

US-09-917-376-1 (1-957) x US-09-949-016-16031 (1-19383)

|  |          |   |       |
|--|----------|---|-------|
| Qy   | 783      | AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro    | 802   |
| Db   | 11492    | TCTCCATCACCATTACCATCACCATCACCATCACCATCACCATCACCATCACCA          | 11493 |
| Qy   | 803      | SerProSerProSerSerProSerProSerProSerProSerProSerProSerSer       | 822   |
| Db   | 11432    | TCGCCATCACCATTACCATCACCATTCCATCACCATCACCATCACCATCTCCA           | 11373 |
| Qy   | 823      | SerProSerProSerProSerProSerProSerProSerArgSerProSerProSerAla    | 842   |
| Db   | 11372    | TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCTCCA          | 11313 |
| Qy   | 843      | SerProSerProSerSerProSerProSerProSerSerProSerSerProSerProSerPro | 862   |
| Db   | 11312    | TCACCATCACCATTCCATCTCCATCACCATCACCATCTCCATCACCATCACCATCTCCA     | 11253 |
| Qy   | 863      | ThrProSerSerSerPro  | 868   |
| Db   | 11252    | TCACCATCACCATCTCCA  | 11235 |
| RESULT 4   |          |   |       |
| US-09-318-448-11/c   |          |   |       |
| ; Sequence 11, Application US/09318448                                 |          |   |       |
| ; Patent No. 6210950   |          |   |       |
| ; GENERAL INFORMATION:   |          |   |       |
| ; APPLICANT: Johnson, William G.                                       |          |   |       |
| ; APPLICANT: Stenroos, Edward S.                                       |          |   |       |
| ; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING |          |   |       |
| ; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS                          |          |   |       |
| ; FILE REFERENCE: 601-1-057  |          |   |       |
| ; CURRENT APPLICATION NUMBER: US/09/318,448                            |          |   |       |
| ; CURRENT FILING DATE: 1999-05-25                                      |          |   |       |
| ; NUMBER OF SEQ ID NOS: 46   |          |   |       |
| ; SOFTWARE: Patent In Ver. 2.0   |          |   |       |
| ; SEQ ID NO 11   |          |   |       |
| ; LENGTH: 18596  |          |   |       |
| ; TYPE: DNA  |          |   |       |
| ; ORGANISM: Homo sapiens   |          |   |       |
| US-09-318-448-11   |          |   |       |
| Alignment Scores:  |          |   |       |
| Pred. No.:   | 1.42e-07 | Length:   | 18596 |
| Score:   | 345.00   | Matches:  | 71    |
| Percent Similarity:  | 83.9%    | Conservative:   | 2     |
| Best Local Similarity:   | 81.6%    | Mismatches:   | 13    |
| Query Match:   | 6.7%     | Indels:   | 1     |
| DB:  | 3        | Gaps:   | 0     |
| US-09-917-376-1 (1-957) x US-09-318-448-11 (1-18596)                   |          |   |       |
| Qy   | 783      | AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro    | 802   |
| Db   | 10388    | TCACCATCACCATTCCATCTCCATCACCATCACCATCACCATCACCATCACCATCACA      | 10329 |
| Qy   | 803      | SerProSerProSerSerProSerProSerProSerProSerProSerProSerSer       | 822   |
| Db   | 10328    | TCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACCATCTCCATCACA     | 10269 |
| Qy   | 823      | SerProSerProSerProSerProSerProSerProSerArgSerProSerProSerAla    | 842   |
| Db   | 10268    | TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACA     | 10209 |
| Qy   | 843      | SerProSerProSer-SerSerProSerProSerSerProSerProSerProSerPr       | 862   |
| Db   | 10208    | TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCTCCATCACCATCACC     | 10149 |
| Qy   | 862      | oThrProSerSerPro  | 868   |
| Db   | 10148    | ATCTCATCACCATCACA   | 10130 |
| RESULT 5   |          |   |       |
| US-09-577-266-11/c   |          |   |       |



```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 701, 13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc_feature
; LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
; LOCATION: 5062, 5167, 11069, 13298, 14479, 14730, 14796, 15344, 15450,
; LOCATION: 15503, 15590, 15840, 16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc_feature
; LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
; LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
; LOCATION: 15042, 15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc_feature
; LOCATION: 1322, 1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc_feature
; LOCATION: 2594, 11293, 16199, 16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc_feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc_feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-665-8

Alignment Scores:
Pred. No.: 1,42e-07 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.9% Conservative: 2
Best Local Similarity: 81.6% Mismatches: 13
Query Match: 6.7% Indels: 1
DB: 3 Gaps: 0

US-09-917-376-1 (1-957) x US-09-962-665-8 (1-18597)
Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerIeuSerPro 802
::: |||||
Db 10388 TCACCATCACCATTCTCCATCTCCATCACCATCACCATCACCATCACCATCACCACCA 10359
Qy 803 SerProSerProSerSerSerProSerProSerProSerProSerSerSerProSerSer 822
::: |||||
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCACCATCTCCATCACCATCTCCATCACCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerProSerAla 842
::: |||||
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACCACCA 10209
Qy 843 SerProSerProSerSerSerProSerProSerProSerProSerProSerProSerProSerPr 862
::: |||||
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCACCATCTCCATCACCATCACCACCA 10149
Qy 862 oThrProSerSerPro 868
::: |||||
Db 10148 ATCTCCATCACCATCACCACCA 10130

RESULT 7
US-09-963-333-8/c
; Sequence 8, Application US/09963333
; Patent No. 6664062
; GENERAL INFORMATION:
; APPLICANT: Scanton, Jr., Vincent P.
; TITLE OF INVENTION: THYMIDINE SYNTHASE GENE SEQUENCE VARIANCES
; TITLE OF INVENTION: HAVING UTILITY IN DETERMINING THE TREATMENT
; TITLE OF INVENTION: OF DISEASE
; FILE REFERENCE: 11926-015002

```

```

; CURRENT APPLICATION NUMBER: US/09/963,333
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 18597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 701..13751
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc_feature
; LOCATION: 716..1293
; LOCATION: 5082..5167, 11069..13298, 14779..14730, 14796..15344, 15450..
; LOCATION: 15503..15590, 15840..16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc_feature
; LOCATION: 732..1379, 1590..2488, 3212..5006, 11238..11422, 11586..
; LOCATION: 12598..13171, 13645..13782, 13806..13813, 14586..14788..
; LOCATION: 15042..15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc_feature
; LOCATION: 1322..1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc_feature
; LOCATION: 2594..11293, 16199..16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc_feature
; LOCATION: 3619..
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc_feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-963-333-8

Alignment Scores:
Pred. No.: 1..42e-07 Length: 18597
Score: 345.00 Matches: 71
Percent Similarity: 83.9% Conservative: 13
Best Local Similarity: 81.6% Mismatches: 1
Query Match: 6.7% Indels: 1
DB: Gaps: 0

US-09-917-376-1 (1-957) x US-09-963-333-8 (1-18597)

Qy 783 AlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerPro 802
Db :::|||||
Dy 10388 TCACCATCACCATTCTCCATCTCCATCACCATCACCATCACCATCACCA 10329
Qy 803 SerProSerProSerSerProSerProSerProSerProSerProSerProSer 822
Db 10328 TCACCATCTCCATCACCATCACCATCACCATCTCCATCACCATCACCATCACCA 10269
Qy 823 SerProSerProSerProSerProSerProSerProSerProSerArgSerProSer 842
Db 10268 TCACCATCACCATCACCATCTCCATCACCATCACCATCACCATCTCCATCACCA 10209
Qy 843 SerProSerProSer-SerSerProSerProSerSerProSerProSerProSerPr 862
Db 10208 TCACCATCACCATCACCATCACCATCACCATCACCATCTCCATCTCCATCACCA 10149
Qy 862 oThrProSerSerPro 868

```



```

DB: 3 Gaps: 56
US-09-917-376-1 (1-957) x US-09-136-574A-2 (1-6416)
QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProile 37
Db 1189 TCTGTAGTGCAGAACAA-----GCTGCATCAGCTTCGGGTTCAGTT 1233
QY 38 AlaIleThrAlaSerProAlaHisAlaAla----- 47
Db 1234 GTAATAAGAAAGAAAGAAATCTCAGAAAGCAGCTTCTTATCTCCAACATGCCAAGACCTG 1293
QY 48 -----ThrThrGlnPro-----TyrThrTrpSerAsnValAlaIle 59
Db 1294 TTTGAATTTGCCGATACACCAAGAAGTGATCGGGGTATATCTGCTGCAACAGGTTCCTAC 1353
QY 60 GlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyr 79
Db 1354 ACATCGGGTGGTTTATGTATGACCTTGGATGG-----GCTGCTGTATGGCTTTAT 1404
QY 80 ValArgThrAsp----- 84
Db 1405 ATTGCGCAAAATCAGACAGTAGTTATTTCAGCAAGCTGAAGAGTTGATGTCAGATATGCT 1464
QY 85 GlyGlyMetTyrArgTrp-----AspAlaIleAsnGlyArgTrpIlePro 99
Db 1465 AATGGTACTAATAACATGACACCAATGCTGGATGATGTTCCATATGCAACATTCATCATG 1524
QY 100 LeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 1525 CTTGCCAAGATTACAGGGAAGAGTTA-----TATAAAGGAGCTGTGGAAAGAACTTA 1578
QY 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMet---TyrThrAsnSerTrp 138
Db 1579 GACCATGGACTACAGAAATTAGTATACCGCAAGAGGATGGCATATCTGCACAGATGG 1638
QY 139 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly----- 152
Db 1639 -----GGTTCAATTAAGATATGCGCAACACAGCTGCATTTTATGATGTGCTAT 1686
QY 153 AlaThrTrpGln----- 158
Db 1687 GCAGACTGTGTCAGGGTCCGATTCCGAACAAAGACCAAAATATTGAACTTTGCAAAAGC 1746
QY 159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArg-----GlyMetGly 174
Db 1747 CAGATGACTGATCCACTGGGTCCACA-----GGTAGAAGTTTGTAGTAGGATTTGGC 1800
QY 175 GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194
Db 1801 ACCAATTATCCACACATCCGCATCAGAGAT----- 1833
QY 195 LysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePro 214
Db 1834 -----GGCAGATGTTTCATGGCTAACGATGAAAAATACCA 1869
QY 215 Asp-----ValGlyThrTyrIleAlaAsnProThrAspThrThrGly 228
Db 1870 GAGTATACAGACACATATTATATGAGCAGCTGTTGGTGGTCTCTGTAGTAGTATAGT 1929
QY 229 TyrGlnSerAspIleGlnGlyValValTrp-----ValAlaPheAspLysSerSer 246
Db 1930 TATAATCATGACATTACCGATTATGACAAAATGAGGTTCCTGCCGATTAATATGCTGGA 1989
QY 247 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 266
Db 1990 ATTGTTGGTGCACTGGCAAG----- 2010
QY 267 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286
Db 2011 -----ATGTACCAGTTATATGAGGTGAACCTATTGAT--- 2043
QY 287 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThr----- 304

```

---

```

Db 2044 ---GATTTTAAAGCAATTTGAAACACCCCAAAATGATGAAATTTTGTTCATCAAAATTT 2100
QY 305 SerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThr 324
Db 2101 GGGAAATTCACAGGTTCAC-----AATTATACCGAAGTAATTTCTTATATCTATAAT 2151
QY 325 SerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsn---AspTyr 343
Db 2152 CGAACAGGATCG-----CCACCAAGGGTAACCTGATAAACTAAGTTTAAATAT 2199
QY 344 Phe-----GlyTyrSerGlyLeuThrIleAspArgGln 354
Db 2200 TTTATAGACCTTAACCGAATTAATCCAGCAGCGGTATTCG----- 2238
QY 355 HisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePhe 374
Db 2239 ---CCTGATGTTGTCAAAGTTGACACA-----TAC 2265
QY 375 ArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsn 394
Db 2266 TACATCGAAGGAGGTAAATAGCGGTCTTACGTATGGAC-----AAAAAT 2313
QY 395 ArgSerLeuArgTyrValLeu---AspIleSerAlaGluProTrpLeuThrPheGlyVal 413
Db 2314 AGGAATATATATCTATGTTCTTGTGATTTAGTGGAACCAAG-----ATA 2358
QY 414 GlnProAsnProProValProSerProLysLeuGlyTrp-----MetAspGlu 429
Db 2359 TATCCTCGCGGTGAAGTTGAACACAAAGAGCGCTCAATTTAAATATCTGTCGCGAG 2418
QY 430 AlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeu 449
Db 2419 GGGTATCCATGGATCTTACCAAT---GATCCTCATATAGGATTAACAGTCAATTA 2475
QY 450 TyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMet 469
Db 2476 GAAAGAATAAATAATATTGTCGCATATGATAATAATAAT-----CTG 2517
QY 470 ValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPro----- 484
Db 2518 GTATGGGGTTTTAGAGCCGGGTGCGGCACATCCACCTGCCAACATCAACACCAACA 2577
QY 485 -----ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 502
Db 2578 CCAACCCCGCCCCCAACACCAACAGTGACAGCA----- 2610
QY 503 AlaAspValThrAlaValProSerThrIlePheThr-----SerProValPheThr 519
Db 2611 -----ACGCCGACCCGACTCTTACACCGACCCGCGGGTCACT---GGT 2655
QY 520 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 539
Db 2656 ACGGAAAGTGGTGTGAAGGTACTGTACAAAGAACAAATGAGCAAGTGCAGACAGCTTCT 2715
QY 540 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 559
Db 2716 ATAAGGCCG----- 2724
QY 560 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579
Db 2725 ---TGGTTTAAG-----ATAGTGAATGGAGGC-----AGCAGCAGT 2757
QY 580 AlaAspGlySerArgPhe-----ValTrp-----AlaProGlyAspProGlyGln 594
Db 2758 GTTGATCTTAGCAGGGTTAAGATAAGTACTGGTACACAGTGGATGGTGACAGCCACAG 2817
QY 595 ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAla 614
Db 2818 AGTCGCGTATGT-----CACTGG----- 2835
QY 615 AsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGly 634

```

```
Db 2836 ---GCACAGATAGGGGCAAGCAATGTGACA---TTCAATTTTGTGAAGCTTAGCAGCGGA 2889
Qy 635 ThrPheTyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGlyLeuPro 654
Db 2890 GTG-----AGTGGAGCGGATTATTAC---CTGGAGGTAGGATTT--- 2925
Qy 655 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlu---GlyAspLeu 673
Db 2926 ACAGTGGAGCTGGCGAGTTGAG-----CTGGTAAGGACACAGGGGATATA 2973
Qy 674 TrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla 693
Db 2974 CAGGTAAGGTTTAAACAGATCACTGGAGCAATACATCAGGCAGACACTGGTTCATGG 3033
Qy 694 IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
Db 3034 TTGCAGAGCATGACCAAT-----TATGAGAGAAATGCCAAGGTGACGCTG 3078
Qy 714 TyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAsp 733
Db 3079 TAT-----GTAGATGGT----- 3090
Qy 734 AspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGly 753
Db 3091 -----GTTCGTGTA-----TGGGGG 3105
Qy 754 GlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 773
Db 3106 CAG----- 3108
Qy 774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSer 793
Db 3109 -----GAGCGGAGGAGCG----- 3123
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerSerProSerProSer 813
Db 3124 -----ACACTGCACCGCAACAGCACACCAACCAACG 3156
Qy 814 ProSerProSerSerProSerSerProSerProSerProSerProSerProSerProSer 833
Db 3157 CCAACTCGACAGCAACCCCAACACTACACTACACCGCCCGACACCGACAGTGAGT 3216
Qy 834 ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerProSer 853
Db 3217 GCAACGCCAACACACCGCACCGCATCACCGTAGTGGCAGTTACTGGACGCCAGT 3276
Qy 854 SerSerProSerSerProSerProSerProSerProSerProSerProSerGlyVal 873
Db 3277 GAGAGT-----TACGGTGGCGCTG 3294
Qy 874 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 893
Db 3295 AAGGTATGGTATGCGAATGGGAATTTAAGACGCCGACGAATGTAATGAATCCTAAGATA 3354
Qy 894 GlnValAlaSerThrGlySerSerSerValAspLeuSerThrValThrValArgTyrTrp 913
Db 3355 AGATAGAGATGTTGGGACGACGACGCGTAGATCTTACAGGTTGAAGTAGACTGG 3414
Qy 914 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaIleGly 933
Db 3415 TACACCATAGATGGTGGAGCGACACAGAGTGTA----- 3447
Qy 934 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 947
Db 3448 -----AGTGTAGCGAGCAGCATAAATCTCGC 3474
```

## RESULT 10

```
US-09-252-991A-13656
; Sequence 13656, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13656
; LENGTH: 8211
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13656
```

## Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 1,04e-06 | Length:       | 8211 |
| Score:                 | 319.50   | Matches:      | 290  |
| Percent Similarity:    | 32.6%    | Conservative: | 133  |
| Best Local Similarity: | 22.3%    | Mismatches:   | 449  |
| Query Match:           | 6.2%     | Indels:       | 432  |
| DB:                    | 3        | Gaps:         | 69   |

US-09-917-376-1 (1-957) x US-09-252-991A-13656 (1-8211)

```
Qy 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 3326 GCAATGGCAGCAGCTCAGCGCAGCTGCGGAACCGGCGAGCAGCGGTGATCTCACCGACG 3385
Qy 38 AlaIleThr-----AlaSerProAlaHisAlaAlaThrThrGlnPro----- 51
Db 3386 GCAACGGCAATCCGATCGCGAGGTCAACCGCGAGCGCGAGCGCAACTGACACCC 3445
Qy 52 -----TyrThrTrpSerAsnVal-----AlaIleGly 60
Db 3446 GTCCACGCCGATCCCAACGGTACTGTGTC--AACGTGGTGGCCGAGACGCCCGCGT 3504
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 3505 AACACAGCCCCCGCGGAGCGGTGACCGTCGATTCAGCGCGCGCGCGCGCGGTGATC 3564
Qy 81 ArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 3565 AACCCGAGCAACGGCGTGTCTATCAGCGCACCGCGCGCGCGGTGCCACCGTGACCTC 3624
Qy 101 LeuAspTrpValGlyTrpAsn-----AsnTrp 109
Db 3625 ACCGAT---GCCGCGGCAACCCGATAGGGCAGGTCAACCGCGACGCGCGCAACTGG 3681
Qy 110 GlyTyr-----AsnGly-----ValValSerIleAlaAlaAsp 120
Db 3682 AGCTTCACGCCGGGACGCCCGCGGCGCAACGCGATGTCGCCACGCCCGACCGAC 3741
Qy 121 Profile-----AsnThrAsnLysValTrpAla-----AlaValGlyMetTyrThr 135
Db 3742 CCGACCGGCAATACCGCGCGCGGCGCGCACCGGTGACCGGTGGCGCGCGCGCGCG 3801
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeu----- 146
Db 3802 CCGGTGATCGATCCGAGCAACCGCACGACCATCAGCGGCAACCGCGAGCGCGGCGCAAG 3861
Qy 147 -----ArgSerSerAspGlnGlyAla 153
Db 3862 GTGATCTCACCGACGGCAACCGCAACCGATCGCGGAACACCCACCGCGACGCGGCG 3921
Qy 154 ThrTrpGlnIleThrPro-----LeuProPheLysLeuGly----- 165
Db 3922 AACTGACCTTCACCGCGCGCGCGCGTGGCGCAACGCGACCGGTGTCTACGCCGTGGCC 3981
Qy 166 -----GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPro 181
Db 3982 CAGGACCTTGGCGGCAATACCGCGCGCGCGCGAGGCAGCACTACCGTGGACGCGGTGGCGCG 4041
```

Qy 182 Asn-----AsnAspAsnIleLeuTyrPheGlyAlaProSerGly 194  
Db 4042 AACAGCGCTGTGGTCAATCCGAGCAACGGCAACCTGCTCAACGGTACCGCGAGCCGGC 4101  
Qy 195 LysGlyLeu-----TrpArgSerThrAspSer 203  
Db 4102 AGCACCCTGACCTTGACCGAGCGCAACGGCAACCGGATCGGCCAGACACCGCGCATGGC 4161  
Qy 204 GlyAlaThrTrp-----SerGlnMetThrAsnPheProAspValGlyThr 218  
Db 4162 AGCGGCAACTGGAGCTTCACGCCCGCTCGCACTACCAAC-----GGCACC 4209  
Qy 219 TyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValVal 237  
Db 4210 GTGGTCAACGTGACCGGAGCGCGCGCAATACCAAC-----GCTCCCGCT 4260  
Qy 238 TrpValAlaPheAspLysSerSerSerLeuGlyGln-----AlaSerLysThr 254  
Db 4261 ACCAGGAGGTGGATTCCTGCTCGCTCGATCCCGGAGGTGGATCCGAGCAACGGTTGC 4320  
Qy 255 IlePheValGlyValAlaAspProAsnProValPhe----- 267  
Db 4321 GTGATCAGCGCACCGCGGACCGCGCAACACCATCATCATCCGATGGCAACGGCAAC 4380  
Qy 268 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 283  
Db 4381 CCGATTGGCCAGGTACCGCGGACCGGAGCGGTAACTGTTCTTCACTCCAGGCATCCG 4440  
Qy 283 ----- 283  
Db 4441 CTGCCGATGGCACGGTGTCAACGTGTGGCGCGCACGCCCAAGCAATGTCACAGTGGC 4500  
Qy 284 -----ThrGlyPheIleProHisLysGlyValPheAspProValAsn 297  
Db 4501 CCGCGGTGTATCACTGTGGATGGCGTGGCCCGCGCGCGCGGTGATCGATCCAGCAAC 4560  
Qy 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr-----AspGlySer 314  
Db 4561 GGCACCGAGATAAGCGGTACCGCGGAGCGCGCGCGGATGCTCTCACCGATGGCGGC 4620  
Qy 315 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrpTrpArgIleSerPro 333  
Db 4621 GGC AACCCGATCGGCCAGCGCACCGCGGACCGGCACTGGACGTTCAACCCCGCGC 4680  
Qy 334 ValPro-----SerThrAspThrAlaAsnAspTyr 343  
Db 4681 ACCCGCTGGCAACGGCACCGGTATCAACCGCTGGCGCGACCGCGCGCAATACC 4740  
Qy 344 PheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro----- 356  
Db 4741 AGCGTCCGCGCAGCGTCACGTCGATGCCATCGCCCGCGCGCGGTGATCAATCCG 4800  
Qy 357 ---AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArg 375  
Db 4801 AGCAACCGAGTGTGTCATCAGCGGTACCGGGAAGCCGCGGCACTGGACGTTCAACCCCGCGC 4854  
Qy 376 SerThrAspGlyGly-----AlaThrTrpThrArgIleTrpAspTrp 389  
Db 4855 ---ACCAGCGCAACCGGCAACCCGATCGGCCAGGTACACCGCGGACCGCGCAAGTGG 4911  
Qy 390 Thr-SerTyr-ProAsnArg-----SerLeuArgTyrValLeuAs 402  
Db 4912 GCTTTCAGCCCGCACCGCGTTGGCCATGCGACGGGTGATCAATGCGTGG----- 4963  
Qy 402 pIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValProSerPr 422  
Db 4964 ---CCAGGACGCGCGCGCAACACAGCAGTC-----CCACCGCGCA---CC 5007  
Qy 422 oLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLe 442  
Db 5008 GTCAGTCTGCTGGCGCCAGCA-GCCCGCGGTGATCGATCCGAGCAACCGGTAGC---GTGAT 5063  
Qy 442 uTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAs 459

Db 5064 CCCTGGTACCCCGAGGCTGGTGCACCGTGTATCTCACC-----GA 5105  
Qy 459 pSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAs 479  
Db 5106 CGGCAACGGCAAC-----CCGATCGGCAGGTACCGCGCGATGGCAGCGCGCA 5153  
Qy 479 nAspLeuIleSerProSerGlyAlaProLeu-----IleSerAlaLe 494  
Db 5154 CTGGAGCTTACGCC-----GGCACCGCTGTCCAATGGCACCGTGTCAATCGGT 5207  
Qy 494 uGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaValProSerThrI 512  
Db 5208 GGCCAGGACGCTCGCGCAACACCGAGCGCGCGCGCACCGACCGATCGGTGGTGC 5267  
Qy 512 ePheThrSerProValPhe-----ThrThrGlyThrSerVal---AspTyrAlaGluLe 529  
Db 5268 CGCGCGCGCGCGGTGATCGACCGGCAACCGCGATGATCGCGGTACCGCGGAAGC 5327  
Qy 529 uAsnProSerIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspAr 549  
Db 5328 CGGTGCGAGGTGATCTCTCACCGATGGCGGCAACCGGATCGGCCAGGCC----- 5379  
Qy 549 gHisValAlaPheSerThrAspGlyGlyLysAsnTrp----- 561  
Db 5380 -----ACCGCGATGGCAGCGCAACTGGAGCTTCAACCCCGGCGCACCGCGCT 5426  
Qy 562 -----PheGlnGlySerLeuProGlyGlyValThrThrGlyG 574  
Db 5427 GGCACACGGCACGGTGTCAATCGGTGGCGCGCAGGATCCGCGCGCAATACCGCGGCC 5486  
Qy 574 yThr----- 575  
Db 5487 GACCAGCACACGGTGGAGCGGTGGCGCGCGCGCGCGGTGTCAACCCCGAGCAACCG 5546  
Qy 576 -----ValAlaLaserAlaAspGlySerArgPheValTrpAlaProGlyAspProG 593  
Db 5547 CAGCGTGTATCGCTACCGCGGAAGCGCGCGCGCGCGGTGATCTCACCGAGCGCGCG 5606  
Qy 593 yGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnG 611  
Db 5607 CAACCCGATCGCGCGAGGTCAACCGCGCGCGCGCGCGCAAC---TGGAGCTTCAACCGCG 5663  
Qy 611 yValPro---AlaAsnAlaGlnIleArgSer----- 620  
Db 5664 CACCGCGTGGCCCAACCGCTCGGTGATCAATCGGTGGCGCGCGCGCGCGCAACAC 5723  
Qy 621 -----AspArgValAsnProLysThr---PheTyrAlaLe 631  
Db 5724 CAGCGCGCGCGCGCACCGACCGGTGACTCGGTAGCCCGCGCGCGCGCGGTCTCGATCC 5783  
Qy 631 uSerAsnGlyThrPheTyrArgSer-----ThrAs 641  
Db 5784 GAGCAACCGTACCGGTGATCAGCGGTACCGCGCAAGCCCGCGCGCGCGGTATCTCACCGA 5843  
Qy 641 pGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyV 661  
Db 5844 CGCGCGC-----GGCAACCGGATACCGGACCGCACCGCGCGCGCGCGCAACTGGAG 5897  
Qy 661 alMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGlyL 681  
Db 5998 CTTCACTCCGGGCACA-CCGCTGACCAACCGGACCGGTGATCAATCGGTGGCGCGGAGCAG 5956  
Qy 681 euTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaIleThrGlyValSerS 699  
Db 5957 CCGCGCGCAACACCGAGCGGTCCGCTCAGCACCAACAGTGGACGCGGTGGCGCGCGCGCACCC 6016  
Qy 599 erAlaValAsnValGlyPheGlyLysSerAlaProGlyLysSerSerTyrProAlaValPheV 719  
Db 6017 CGGTGATCGACCGGCAATGTTGTCAAACTCAGCGCGCACCGCGCGCGCGCGCGCGCG 6076  
Qy 719 alValGlyThrIleGlyGly----- 725

```

Db 6077 TGATCTCACCAGTGGCAATGCAACCCGATCGGCCAGACCTCGCCGACGGTAGCGCA 6136
Qy 726 -----valThrGlyAlaTyrArgS 732
Db 6137 ACTGACCTTACACCGCGGACCGCTGGCCAAAGCAGCGTGTCAACGCGGTGGCC 6196
Qy 732 eAspAspCysGlyThrThrTrpValLeuLeuAsnAspGlnHisGlnTyrGlyAsn 752
Db 6197 AGGACCGCGCGCAATACAG-----CGTCCGGCAGCAGCAGCAGCGTGTGATACG 6246
Qy 752 rpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-ValTyrIleGlyThrAsn 771
Db 6247 GTGGCGCGCGCACCGCGTGTATCAATGCCAAGCAGCGAGCGTGTATCACCGCACCGCC 6306
Qy 772 -----GlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGlySer-ProSe 789
Db 6307 GAGGTGGCGCGCAAAAGTGTCTCACCGACGCAACCGCATCGCGGAGACCAACC 6366
Qy 789 rProSerVal-----SerProSerAlaSer-----ProSerLeu----- 800
Db 6367 GCCGACGCGAGTGGCAACTGACCTTACCCCGCGCAGCGCGCTGGCCCAACGTCAGGTG 6426
Qy 801 ----SerProSerPro-----SerProSerSerProSerProSerProSerPro----- 816
Db 6427 ATCAACGCGTGGCGGAAGACCGCGCGGCAACGCGCGTGTCCGCGCACCAACGCGTG 6486
Qy 817 -----SerSerSerProSerSerSerProSerPr 826
Db 6487 GACTCGTGGCGCGTCCGCTCCGCTCGCTGAGCATCAGCGCGGCGCGCTGCTGACC 6546
Qy 826 oSerPro-----SerProSerProSerProSerAr 836
Db 6547 GGCACCGCGGAGCGGACAGTCAGGTGCGTATCGTGTGTCACGCGGACACCGCAACCGC 6606
Qy 836 gSerProSerProSerAlaSerProSerProSerSer-----Se 849
Db 6607 ATCAGGTACCGTCCGCGGCGCGGCAACTTCAGCTGCGCGCGCGCGCGTGCATC 6666
Qy 849 rProSer-----ProSerSerProSerSerSe 859
Db 6667 ACCGCGCACTGATCGCGCGGTTGCCGTGAGCGCGCGCGCAAGTCAGCGCGCGCGCC 6726
Qy 859 rProSer-----ProThrPr 864
Db 6727 ACCATCAACGCGCGGACCTGGCGCGCGCGACCATCAGCGTGGCGGAAGCGCGGATACC 6786
Qy 864 oSerSerProValSerGlyValLysValGlnTyrLysAsnAsnAspSerAlaPr 884
Db 6787 TGGATCAACGCGCGGAGATCGGGACGCGC-ATCCAGTGC----- 6825
Qy 884 oGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerValAs 904
Db 6826 ----GATGTGACGTTCGCTCCGACCATCAGCTCGC----- 6858
Qy 904 pLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuVa 924
Db 6859 ----CAGTGTGTACGCTCAAGTTC---GCCGGCAGAACGGCTACAGGCGCGAGGTGAG 6911
Qy 924 lTyAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPhe----- 941
Db 6912 CCATACCTTACCCCGCGGACATCGCGCGGCAACCTGACCTGACCTGACGCGCTCC 6971
Qy 942 -GlySerValAsnPro-----AlaThrProThrAlaAsp 952
Db 6972 CGCGGATGCGCGCGTTCGCGGAGGCGCGCTCGACGCTCACCGCGGAC 7020

```

## RESULT 11

```

US-09-410-551B-28
; Sequence 28, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER

```

```

; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTIL, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4766)
; US-09-410-551B-28

```

```

Alignment Scores:
Pred. No.: 7,82e-06 Length: 4767
Score: 296.50 Matches: 233
Percent Similarity: 35.2% Conservative: 101
Best Local Similarity: 24.5% Mismatches: 397
Query Match: 5.8% Indels: 223
DB: 3 Gaps: 46

```

US-09-917-376-1 (1-957) x US-09-410-551B-28 (1-4767)

```

Qy 9 LeuThrMetArgSerArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaVal 28
Db 1827 CTGACGGAACACGAAGACCGCTGCGCGCTACCTACCTGCGCGCGCGGATATA 1886
Qy 29 AlaAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 48
Db 1887 CGGGCTGTGGCATCCACGCTG-----GCGTGACACGCTCGGT-GTTCGAGCACCGCGC 1939
Qy 49 ThrGlnProTyr----- 52
Db 1940 CGTACTCTCTGGAGATGACACCGTCACCGGACCGCGGTGACCGACCCAGGATCGTGT 1999
Qy 53 -----ThrTrpSerAsnValAlaIleGlyGlyGly-----PheVal 65
Db 2000 TGTCTTTCCGCGAGGGGTGGCAGTGGCTGGGGATGGGCGAGTGCACCTGCGCGATTTCGTC 2059
Qy 66 AspGlyIleValPheAsnGly-----AlaProGlyIleLeuTyrValArgThrAsp 83
Db 2060 GTGTGTGTTCGCGGAGCGGATGCGCGGCGGTGCGCGCGGTTCGCGAGTTCGT---GGA 2116
Qy 84 lIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 2117 CTGGGATCTGTTCAGGTTCGTGATGATCCCGCGGTGG-TGGACCGGTGTATG---TGG 2172
Qy 104 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123
Db 2173 TCCAGCCC-GCTTCTCTGGCGATG-----ATGTTTCTCTGCGCGC----- 2213
Qy 124 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSer---TrpAspProAsnAsp 142
Db 2214 -----GTGTGCGAGCGCGCGGTGTGCGCGCGGATGCGGTGTATCGGCCATTCGCAG 2264
Qy 143 GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGln----- 156

```

```
Db 2265 GGTGAGATCGCGCAGCTTGTTGTGGCGGCTCGGTGTCACTACGCGATCGCGCCCGGATC 2324
Qy 157 IleThrProLeuProPheLeuLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176
Db 2325 GTGACCTTGCGCAGCAGCAGCGGCTCGCGGGCGCTGGCGGGCGCGGCGGATGGCATCC 2384
Qy 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
Db 2385 GTGCGCCCTG---CCCGCGAGAGTGTGAGCTGTGTCAGCGGGCC----- 2426
Qy 197 LeuTyrArgSerThrAspSerGlyAlaThrTyrSerGlnMetThrAsnPheProAspVal 216
Db 2427 ---TGGATCGCGGCCCAACAGCGGCCCTCCACCGTGTACCGGCGACCCCGGAAGCG 2483
Qy 217 GlyThrTyrIleAlaAsnProThrAspThrThrGly----- 228
Db 2484 GTCGACCATCTCTCACCGCTCATGAGCGCACAGGGGTGCGGTGCGCGGATCACCGCTC 2543
Qy 229 ---TyrGlnSerAspIleGlnGlyValValTyrPheVal-----AlaPheAspLysSer 244
Db 2544 GACTATCGCTCGCACACCCCGCAGCTGAGCTGATCGCGCAGCAACTACTCGACATCACT 2603
Qy 245 SerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsn 264
Db 2604 AGCAGACAGCTCGCAGACCCCGCTCGTCCGCTGGCTGTCGACCGTGGACGGCACC--- 2660
Qy 265 ProValPheTrp-----SerArgAspGlyGlyAlaThrTyrGlnAlaValProGlyAla 282
Db 2661 -----TGGTTCGACAGCCCGCTGGACGGGAGTACTGGTACCAGGAACCTCGCGTAA 2711
Qy 283 ProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 302
Db 2712 CCGGTGCGTTTCCACCCCGCGCTGAGCCAGTGTGAGCGCCAGCGGCGCACCGTGTTCGTC 2771
Qy 303 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTyrLysPheSer 322
Db 2772 GAGGTACGCCCGCCCGCGGTGTTGTCAGCGGATGAGCAGCAGTGTGTCACGGTTGCC 2831
Qy 323 Val-----ThrSerGlyThrTyrTyrThrArgIleSerProValProSerThrAspThr 339
Db 2832 ACGTGCGTGTGACGAGCGGCGGCGCCACCGGATG-----CTCACCGCCCTG 2879
Qy 340 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 359
Db 2880 GCACAGCGCTAT-----GTCCAGCGGCTCACCGTCGAC----- 2912
Qy 360 MetValAlaThrGlnIleSerTyrTyrProAspThrIleIlePheArgSerThrAspGly 379
Db 2913 -----TGCCCGCCATCCTC----- 2927
Qy 380 GlyAlaThrTyrThrArgIleTyrAspTyrThrSerTyrProAsnArgSerLeuAtgTyr 399
Db 2928 GGCACACCAACACCCGGGACTGGACCTTCGACCTTCCGACCTTCCACACACCGCGGTAC 2987
Qy 400 ValLeuAspIleSerAlaGluProTyrLeuThrPheGlyValGlnProAsnProProVal 419
Db 2988 -----TGCTC-----GAGTCGGCTCCCGCGCC 3011
Qy 420 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 439
Db 3012 ACGGCC-----GACTCGGC 3026
Qy 440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAsp 459
Db 3027 CACCCGCTCTCGGCACCGAGTGGCGTGGCGCGG----- 3062
Qy 460 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAla---Val 478
Db 3063 TCGCCGGCGGGTGTTCACGGGTCCCGTCCCGCGCGGTGCGGACCGCGGGTGTTCATC 3122
Qy 479 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498
Db 3123 GCCGAACCTGGCGTGGCGCGCGCGGACGCCACC-----GACTGCGCC 3164
```

---

```
Qy 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518
Db 3165 ACGGTGAAACACAGCTCGAGCTCACCTCGCTGCGCGCGGATCGCGCGCAGGGCC--- 3221
Qy 519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 538
Db 3222 ACCGCGCAGACTCGGTGCTGAT----- 3242
Qy 539 SerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558
Db 3243 -----GNAACCGCGCGCAGCGGGCGCGCTTCACCGTCCACACCGCGCTCGCGAC 3296
Qy 559 LysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAla 578
Db 3297 GCCCGTGGAGCTGCACCGCAG-----GGGGTCTCCGCGCCCGCGCGCTG---CC 3347
Qy 579 SerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnProValValTyr 598
Db 3348 CAGCCCCGAAGCGCTGCACACCGCTGCGCCCG-----CCGGC-----CGGTGCGC 3395
Qy 599 AlaValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618
Db 3396 GCGACCGGCTGCCCGGGCGTGGGACGCGCGGACCGAGGTCTTCGTGAAGCCGAAGTC 3455
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThr 635
Db 3456 GACAGC-----CCTGACGGCTTCTGGGCACACCCCGACCTGCTGCAGCGGTGTC 3503
Qy 636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655
Db 3504 TTCTCCCGGTGCGCGACGGG-----AGCGCCAGCGCAGCGGATG---GCGCGACCTGCG 3556
Qy 656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeu 675
Db 3557 GGTGCACGCTGCGACGCCACCGTGTGCGCGCTGCTCACCCCGCGCAGTGGTGT 3616
Qy 676 AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTyrSerAlaIleThr 695
Db 3617 CGTGGAGCTCGCCCGCTTCGAGCGGTGCGGATGCGGT-----GCTCACCGC 3664
Qy 696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713
Db 3665 GGAGTGGTGAAGCTGGCGGAGGTGCGTTCGCGAGCGGATCCGACGAGTCCGACGGTCT 3724
Qy 714 TyrProAlaValPheValValGlyThrIleGlyGlyVal---ThrGlyAlaTyrArgSerAs 733
Db 3725 GCTTCGCTTGAAGTGTGCGGTGCGGAGGCCACTACGACGCTGCC-----GA 3775
Qy 733 pAspCysGlyThrThrTyrValLeuIleAsn-----AspAspGlnHisGlnTyr 749
Db 3776 CGACTCCCGAGGGGTACACCTTCATCAGCGGCACACACCCCGACGACCCCGCAGCACC 3835
Qy 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal----- 766
Db 3836 CACCAACCCCAACACACACCGCACCGCACACAAACACCAACCGGCTCTCACCGC 3895
Qy 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780
Db 3896 CCTCCAACACCCACTCATCACCACCAACCGCTCATCTGCTCCACACACCGACGCC 3955
Qy 781 ---GlyGlyAlaProSerGly---SerProSerPro-----SerValSerProSerAla 797
Db 3956 CCCAGGCGCGCTGACCGGCTCACCAGCGCTCACCGCACCAAAACGAAACCGCGCGCAT 4015
Qy 797 erProSerLeuSerPro-----SerProSerProSerSerSerProSer 811
Db 4016 CCACCTCATCGAAACCCACACCCACACCCGACCCGCTCCCTCACCCTCACCACCT 4075
Qy 811 erProSerPro-----SerProSerSerSerProSerSerProSerProSer 825
Db 4076 CCACCAACCCACCTACGCTCTCAACCAACACCTCCACACCCCGCCACCTCACCCCAT 4135
```



[illegible]



2928 GGCACCAACACACCGGGTACTGGACCTTCGACCTACGCTTCCACACACGAGCGGTAC 2987  
400 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVal 419  
2988 -----TGGCTC-----GAGTCGGCTCCCGGCC 3011  
420 ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAsp 439  
3012 ACGGCC-----GACTCGGC 3026  
440 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 459  
3027 CACCGCTCTCGCACCGAGTCGCGTCGCGG-----3062  
460 SerGlyGlyGlnIleHisIleAlaProMetValLysLeuGluThrAla----Val 478  
3063 TCGCGGCGCGGTGTACAGGGTCCCGTCCCGCGGTGCGGACCGCGGGTGTTCATC 3122  
479 AsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 498  
3123 GCGGAACCTGGCGTCCCGCGCGCGCACACC-----GACTCGCC 3164  
499 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 518  
3165 ACGTGTGAACAGCTGACGTACCTCGTCCCGCGGATCCCGCGCGAGGCC--- 3221  
519 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 538  
3222 ACCGCGCAGACCTGGGTGAT-----3242  
539 SerPheAspSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 558  
3243 -----GAACCGCGCGCGCGCGCGCGCTTCACCGTCCACACCGCGTCCGCGC 3296  
559 LysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyThrValAlaAla 578  
3297 GCCCGTGAACGCTGACACCCCGAG-----GGGTTCGCGCGCGCGCGGTG---CCC 3347  
579 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 598  
3348 CAGCCGGAAGCGTGCACACCGCTGGCCCG-----CGGGC-----GGGTGCC 3395  
599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618  
3396 GCGGACGGCTCCCGCGCGGTGGCGACGCGCGACGAGTCTTCGTCAAGCGGAGTC 3455  
619 ArgSerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThr 635  
3456 GACAGC-----CCTCAGCGCTTCGTGGCACACCGACCTGCTCGACGCGGTC 3503  
636 PheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 655  
3504 TTCTCGCGGTGGGACCGG-----ACCGCGACCGACCGGATG-CGGGACCTCGC 3556  
656 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 675  
3557 GTGTGACGCGTGGGACCGCACCGTGTGCGCGCTCCCTCACCGCGCGACAGTGTGT 3616  
676 AlaAlaSerSerGlyLeuTyrHisThrAsnGlyGlySerSerTrpSerAlaIleThr 695  
3617 COTGAGCTCGCGCTTCGAGGTGCGGATGCGGATGCGGT-----GCTACCGC 3664  
696 GlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAlaProGlySerSer 713  
3665 GGAGTCGGTGACGTGGGCGAGGTGCGTCCGCGGCGGATCGGACGAGTCGACCGGTCT 3724  
714 TyrProAlaValPheValValGlyThrIleGlyVal-ThrGlyValTyrArgSerAs 733  
3725 GCTTCGGCTTGAGTGTCCCGTGGCGGAGGCCACTACGACGTGCC-----GA 3775  
733 pAspCysGlyThrThrTrpValLeuIleAsn-----AspAspGlnHisGlnTyr 749  
3776 CGAGTGTCCGAGGGGTACACCTCATCCGCGCACACACCGCGACCGCGACCGACCC 3835

QY 749 rGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal----- 766  
Db 3836 CACCAACCCCAACACACACACACACACACACACACAAACACACACCGGTCTCACCGC 3895  
QY 767 -----TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle----- 780  
Db 3896 CTTCAACACACACCTCATCACCAACACACACCTCATCGTCCACACACCGACCC 3955  
QY 781 ----GlyGlyAlaProSerGly-SerProSerPro-----SerValSerProSerAlas 797  
Db 3956 CCAGGCGCGCGCTGACCGGCTTACCGCGCACCGCACCAAAACGACACCCCGCGCAT 4015  
QY 797 erProSerLeuSerPro-----SerProSerProSerSerSerSerPro 811  
Db 4016 CCACCTCATCGAAACCCACACACACACACACCTCCCTCCCTCACTCACTCACTCACT 4075  
QY 811 erProSerPro-----SerProSerProSerProSerProSerProSerPro 825  
Db 4076 CCACCAACCCACCTACGCTTCACCAACACACCTCCACACCTCCCTCACTCACTCACT 4135  
QY 825 erProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 845  
Db 4136 CACCAACCCACACACACACACACACACACACACACACACACCTCACTCACTCACTCACT 4195  
QY 845 erProSerSerSer-----ProSerProSerSerSerSerSerSerSerSerP 860  
Db 4196 CGCATCTCATCACCGCGCGCTCCGCGACCTCCGCGCATCTCCGCGCGCATCTCCGCGCGCATCTCAA 4255  
QY 860 roSerProThrProSerSerSerPro 868  
Db 4256 CCACCCACACCTTACCTCTCTCC 4281  
RESULT 13  
US-09-410-551B-32  
; Sequence 32, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHUSLA, CHAITAN  
; APPLICANT: SANTI, DANIEL  
; APPLICANT: WU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; PRIOR FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 4818  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic  
; OTHER INFORMATION: PKS synthase fragment  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(4817)  
US-09-410-551B-32  
Alignment Scores:  
Pred. No.: 1,18e-05 Length: 4818  
Score: 293.00 Matches: 235  
Percent Similarity: 33.1% Conservative: 97

```
Best Local Similarity: 23.4%  Mismatches: 379
Query Match: 5.7%  Indels: 295
DB: 3  Gaps: 48
```

US-09-917-376-1 (1-957) x US-09-410-551B-32 (1-4818)

|    |      |  |      |
|----|------|--|------|
| Qy | 1    | MetAspArgSerGluAunIleAraGLeuThrMetArgSer-<br>:::                     | 13   |
| Db | 1798 | CTGACCGCTCCCGCGCCGCCGCTCAGCAGCGCGCAGACCTTCCGCTGCTCGTGT<br>:::        | 1857 |
| Qy | 14   | ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly<br>:::     | 33   |
| Db | 1858 | CGGCGGCTCCCGGAGCACTGCAGAGCAGATCGGGCCCTGCGCG-<br>:::                  | 1905 |
| Qy | 34   | ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrThrGlnProTyrThr<br>:::  | 53   |
| Db | 1906 | -----CCTATC-----TCGACACCGCGCGCGCTCGACCGCGCGCTGCGCG<br>:::            | 1950 |
| Qy | 54   | -----TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn<br>:::      | 71   |
| Db | 1951 | AGACACTGG-----<br>:::  | 1959 |
| Qy | 72   | GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr-<br>:::          | 88   |
| Db | 1960 | -----CCGGC-----GTACGCACTTACCACCGGGCGTACTGCTCGGG<br>:::               | 2001 |
| Qy | 88   | -----<br>:::   | 88   |
| Db | 2002 | ACACCGTCATCGCGCTCCCGCGGACCGAGCAGCAACTCGTCTTCTACTCTCCG<br>:::         | 2061 |
| Qy | 89   | -----ArgTrpAspAla<br>:::   | 92   |
| Db | 2062 | GTACGGCACCACGATCCCGCGATGGCGAGCAGCTAGCCGATCGTCTGCTGTGTTCG<br>:::      | 2121 |
| Qy | 93   | AlaAsnGly-<br>:::  | 95   |
| Db | 2122 | CCGAGCGGATGGCGAGTGTGCGCGCGCTTGC CGGAGTTCGTGGAGTGGGATCTGTTC<br>:::    | 2181 |
| Qy | 96   | -----ArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp<br>:::               | 109  |
| Db | 2182 | CGGTTCGGATGATCCCGCGGTGTGACCGGTTGATGTGTGTCCAGCCC-GCTTCTCGG<br>:::     | 2240 |
| Qy | 110  | GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIysValTrpAla<br>:::  | 129  |
| Db | 2241 | CGCATG-----ATGGTTTCCCTGGCGCG-<br>:::                                 | 2273 |
| Qy | 130  | AlaValGlyMetTyrThrAsnSer---TrpAspProAsnAspGlyAlaIleLeuArgSer<br>:::  | 148  |
| Db | 2274 | CGGCGGCTGTGCGCGCGGATCGGTGATCGGCCATTCCGAGGTTGAGATCGCGCGAGCT<br>:::    | 2333 |
| Qy | 149  | SerAspGlnIleAlaThrTrpGln-----IleThrProLeuProPhe<br>:::               | 162  |
| Db | 2334 | TGTTGTGCGGTTGCGGTGTCACTACGCGATGCGCGCGGATCGTGCCTTGCAGCGCCAG<br>:::    | 2393 |
| Qy | 163  | LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn<br>:::  | 182  |
| Db | 2394 | CGCATCCCGGGGCTTGGCGGCGCGGGCGGATGCGATCCGTCGCGCTG-----CCGCG<br>:::     | 2450 |
| Qy | 183  | AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyIysGlyLeuTrpArgSerThrAsp<br>:::  | 202  |
| Db | 2451 | CAGGATGTCGAGCTGTGTCGACGGGCG-<br>:::                                  | 2492 |
| Qy | 203  | SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn<br>:::  | 222  |
| Db | 2493 | AACGGGCGCGCTCCACCGTGATCGCGGGCACCCTCCGGAAGCGGTGCAACCATGTCTCACC<br>::: | 2552 |
| Qy | 223  | ProThrAspThrThrGly-----TyrGlnSerAspIle<br>:::                        | 233  |
| Db | 2553 | GCTCATGAGGCACAGGGGTGCGGGTGC GGATCACCCTGCACATATGCTCGCACACC<br>:::     | 2612 |
| Qy | 234  | GlnGlyValValTrpVal-----AlaPheAspLysSerSerSerLeuGlyGln<br>:::         | 250  |
| Db | 2613 | CCGCACTCGAGTGTATCCCGCAGCACTACTCGATCATCTAGCAGCAGCAGCTCGCAG<br>:::     | 2672 |

|    |      |  |       |                |
|----|------|--|-------|----------------|
| Qy | 251  | AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp         | ----- | 265            |
| Db | 2673 | ACCCCGCTCGTGGCGTGGCTTCACCGTGCAGCGCACC                          | ----- | TGGGTGCAG 2720 |
| Qy | 269  | SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro   | 288   |                |
| Db | 2721 | AGCCCGCTGGACGGGGAGTACTGGTACCGGAACCTGCGTGAACCGGTCTTCCACCCC      | 2780  |                |
| Qy | 289  | HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly   | 308   |                |
| Db | 2781 | GCCGTACAGCAGTGCAGGCCACCGGTGTCGTGAGGTGACGGCAGCCGC               | 2840  |                |
| Qy | 309  | GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal                  | ----- | ThrSer 325     |
| Db | 2841 | GTGTTGTTGAGCGCATGGACGACGATGTCGTACGGTTCACCGTTCGCGTGCAGCAG       | 2900  |                |
| Qy | 326  | GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly   | 345   |                |
| Db | 2901 | GGCGACGCCACCCGGATG-----CTACCGCCCTGGCAGCGCTAT-----              | 2942  |                |
| Qy | 346  | TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle   | 365   |                |
| Db | 2943 | GTCCACGGGTACCGTCGAC-----                                       | 2963  |                |
| Qy | 366  | SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg   | 385   |                |
| Db | 2964 | -----TGGCCCGCATCTC-----GGCACCCACCAACCCGG                       | 2996  |                |
| Qy | 386  | IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla   | 405   |                |
| Db | 2997 | GTACTGACCTTCGACCTACGCCTTCCAACACGACGGGTAC-----                  | 3038  |                |
| Qy | 406  | GluProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGly      | 425   |                |
| Db | 3039 | -----TGGTCTC-----GATCGGTCTCCCGGCCACGGCC-----                   | 3068  |                |
| Qy | 426  | TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr   | 445   |                |
| Db | 3069 | -----GACTCGGGCCACCCGCTCTCGGCACC                                | 3095  |                |
| Qy | 446  | GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHis      | 465   |                |
| Db | 3096 | GGAGTCCCGTCGCCGG-----TCGCGGGCCGGGTTC                           | 3131  |                |
| Qy | 466  | IleAlaProMetValLysGlyLeuGluGluThrAla-----ValAsnAspLeuIleSerPro | 484   |                |
| Db | 3132 | ACGGGTCCCGTCCCGGTGGACGCGCGGTGTTCATCCCGACATGGCGCTGCC            | 3191  |                |
| Qy | 485  | ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp   | 504   |                |
| Db | 3192 | GCGGCCACGCCACC-----GATCGGCCACGGTCGACAGCTGCAC                   | 3233  |                |
| Qy | 505  | ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal   | 524   |                |
| Db | 3234 | GTCACTCCGTGCCCGCGCATCCGCGCGGACGGGCC-----ACCGCGACAGCTGGGTC      | 3290  |                |
| Qy | 525  | AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSer   | 544   |                |
| Db | 3291 | GAT-----GAAACCGCGGCC   | 3305  |                |
| Qy | 545  | GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly   | 564   |                |
| Db | 3306 | GACGGGGCGCGCTTCACCGTCCACCCCGGTGGCGACGCCCGCTGGACGCTGCAC         | 3365  |                |
| Qy | 565  | SerGluProGlyGlyValThrThrThrGlyThrValAlaAlaSerAlaAspGlySerArg   | 584   |                |
| Db | 3366 | GCCGAG-----GGGGTTCCTCCCGCGCGCGGTG-----CCCCAGCCGACCGCTGCAC      | 3416  |                |
| Qy | 585  | PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn   | 604   |                |
| Db | 3417 | ACCGCTGGCCCG-----CGGGC-----GCGGTGCGCGACGCGCTGCCCGG             | 3464  |                |

```

QY 605 SerTIPAlaAalaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624
Db 3465 GCGTGGGAGCCGCGGACGAGTCTCGTGAAGCCGAAGTCGACAGC----- 3512
QY 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641
Db 3513 CCTGACGGCTTCGTGGACACACCCGAGCTGCTCGACGCGGTCTCTCCGGGTCGGGAC 3572
QY 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661
Db 3573 GGG-----AGCGCCGAGCCGCGGATG-CGCGACCTCGCGGTGCACGCGTCGGACGC 3625
QY 662 MetPheHisAlaValProGlyLysGlyGlyAspLeuThrPheLeuAlaAalaSerSerGlyLeu 681
Db 3626 CACCGTCTCGCGCGCTGCTCCACCGCCGCGAGTGTGTGCTGAGTGTGCGCGCTT 3685
QY 682 TyrHisSerThrAsnGlyGlySerSerTIPSerAlaIleThrGlyValSerSerAla--- 700
Db 3686 CGACGGTCCGGATGCCGT-----GCTCACCGGAGTGGTGGTGGG 3733
QY 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719
Db 3734 CGAGGTGCGCTCGGAGCGGATCGACGAGTGGAGCGGTCTGCTCGCTGTGAGTGT 3793
QY 720 ValGlyThrIleGlyVal-ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr 739
Db 3794 GCCGGTGGCGGAGCCCACTACGACGCTGCC-----GACGAGTGGCGGAGGCTA 3844
QY 739 pValLeuIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755
Db 3845 CACCTCATACCGCCACACACCCGAGCAGCCGAGCAGCCCAACCCCAACACAC 3904
QY 755 alIleThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768
Db 3905 ACCACACGACCCACACACACACACGCGTCTCACCGCCCTCCACACACACCTCAT 3964
QY 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe 785
Db 3965 CACCAACACACACACCTCATCTGCTCCACACACACACCCGCGCGCGCTCAC 4024
QY 785 rGly-SerProSerPro-----SerValSerProSerAlaSerProSerLeuSerPro- 802
Db 4025 CGGCTTACCGGACCGCACAAACACACACCCGCGCGCTCATCCTCATCGAAACCCA 4084
QY 803 -----SerProSerProSerSerSerProSerProSerPro----- 814
Db 4085 CCACCCCCACACCCCACTCCCTCTACCCCACTCACCACTCCACCAACCCCACTACG 4144
QY 815 -----SerProSerSerSerSerSerSerProSerProSerProSerProSer 831
Db 4145 CTTACCAACACACACCTCCACACCCCACTCACCACTCACCACTCACCACTCACCA 4204
QY 831 erProSerProSerArgSerProSerProSerAlaSerProSerProSerSerSer- 849
Db 4205 CACCAACACACCCCAACACCCCACTCACCACTCACCACTCACCACTCACCACTCAC 4264
QY 850 -----ProSerProSerSerProSerSerProSerProSerProSerProSer 866
Db 4265 CGGCTCGGACCTCGCGGATCTCGCGCGGATCTCGCGCGGATCTCGCGCGGATCT 4324
QY 866 erSerPro 868
Db 4325 CCTCTCCC 4332

```

RESULT 14

```

US-09-940-316B-32
; Sequence 32, Application US/09940316B
; Patent No. 6759536
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN

```

```

; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE fkbA GENE OF THE FK-520 POLYKETIDE SYNTH
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940,316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 4818
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4817)
US-09-940-316B-32

```

```

Alignment Scores:
Pred. No.: 1.18e-05 Length: 4818
Score: 293.00 Matches: 235
Percent Similarity: 33.1% Conservative: 97
Best Local Similarity: 23.4% Mismatches: 379
Query Match: 5.7% Indels: 295
DB: 3 Gaps: 48

```

```

US-09-917-376-1 (1-957) x US-09-940-316B-32 (1-4818)
QY 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSer----- 13
Db 1798 CTGACCGCTTCCCGGCGCGCGCTCAGCAGCGGCGAAGACCTTCGCTGCTCGTGT 1857
QY 14 ArgArgLeuValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly 33
Db 1858 CGGCGCGTTCCTCCGAGGACACTCGACGAGAGATCGGCGCTGCGCG----- 1905
QY 34 ValLeuProIleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyrThr 53
Db 1906 -----CCTATC-----TCGACACCGCGCGCGCTCGACCGGCGCGTGGCGC 1950
QY 54 -----TrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 1951 AGACACTGG----- 1959
QY 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyr----- 88
Db 1960 -----CCCGC-----GTACGCATTCACCCACCGGCGCTGCTGCGGG 2001
QY 88 ----- 88
Db 2002 ACACCGTCATCGCGCTCCCCCGGCGGACGAGCGGCGGCGGCGGCGGCGGCGG 2061
QY 89 -----ArgTrpAspAla 92
Db 2062 GTCAGGCGACCCAGCATCTCCCGCATGGCGGAGCAGCAGTAGCCGATTCGTGCTGTTG 2121
QY 93 AlaAsnGly----- 95
Db 2122 CCGAGCGGATGGCGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2181
QY 96 -----ArgTIPileProLeuLeuAspTIPValGlyTIPValGlyTIPValGlyTIP 109

```

Db 2182 CGGTTCTGGATGATCCGGCGGTGGTACCGGGTTGATGTGCTCCAGCCC-GCTTCCTGG 2240  
Qy 110 GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAla 129  
Db 2241 GCGATG-----ATGGTTCTCCCTGCGCGG-----GTGTGCAG 2273  
Qy 130 AlaValGlyMetTyrThrAsnSer--TrpAspProAsnAspGlyAlaIleLeuArgSer 148  
Db 2274 GCGCCCGGTGTGCGCGGATCGGTGATCGGCCATTGCGAGGTGAGATCGCGCAGCT 2333  
Qy 149 SerAspGlnGlyAlaThrTrpGln-----IleThrProLeuProPhe 162  
Db 2334 TGTGTGCGGGTGGGTGCTACTACGCGATCCCGCCGGATCGTGACCTTGGCGAGCCAG 2393  
Qy 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsn 182  
Db 2394 GCGATCGCGGGCGCTCGGGCGCGGGCGCGATGCGATCGGTGCGCCCTG---CCGCG 2450  
Qy 183 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp 202  
Db 2451 CAGGATGTCAGCTGGTTCGACGGGGCC-----TGATGCGCGCCAC 2492  
Qy 203 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222  
Db 2493 AACGGCGCGCTCCACCGTATCGGGGACCCCGAAGCGGTGCGACCATGCTCTCACC 2552  
Qy 223 ProThrAspThrThrGly-----TyrGlnSerAspIle 233  
Db 2553 GCTCATGAGGACAAAGGGTGGGGTGGCGGATCACCGTGCAGTATGCTTCGCACACC 2612  
Qy 234 GlnGlyValValTrpVal-----AlaPheAspLysSerSerSerLeuGlyGln 250  
Db 2613 CCGCACGTGCGAGTGTACCGCGAAGTACTCTGACATCACTAGCAGCAGCAGCTGCAG 2672  
Qy 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp----- 268  
Db 2673 ACCCGCTGTCGCGTGGCTGTCACCGTGGACGGCACC-----TGGTGCAC 2720  
Qy 269 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIlePro 288  
Db 2721 AGCCCGTGGACGGGAGTACTGTGTACGGAACTGCTGCGAGTGCAGCGCAGCGCCG 2780  
Qy 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
Db 2781 GCCGTGACGAGTTGACGGCCCGACGACCGGTGTCGAGTGCAGCGCAGCGCCGCG 2840  
Qy 309 GlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal-----ThrSer 325  
Db 2841 GTGTTGTTGACGGCGATGGACGACGATGCTGTCACGGTTGCCACGCTGCGTGCAGCAG 2900  
Qy 326 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 345  
Db 2901 GCGACGCCACCGCGATG-----CTCACCGCTTGGCAGCGCTAT----- 2942  
Qy 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365  
Db 2943 GTCCACGGCGTACCCTGCAC----- 2963  
Qy 366 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 385  
Db 2964 -----TGGCCCGCATCTC-----GGCACCAACCAACCCGG 2996  
Qy 386 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 405  
Db 2997 GTACTGACCTTCGACCTACGCTTCCACACACAGCGGTAC----- 3038  
Qy 406 GluProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGly 425  
Db 3039 -----TGGCTC-----GATGCGCTCCCCCGCCACCGCC 3068  
Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
Db 3069 -----GACTCGGGCCACCCCGTCTCTCGGCACC 3095

Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHis 465  
Db 3096 GGAGTCCCGTCCGCGG-----TCGCCGCGCGGTGTTC 3131  
Qy 466 IleAlaProMetValLysGlyLeuGluThrAla---ValAsnAspLeuIleSerPro 484  
Db 3132 ACGGTTCGCGCGCGCGGTGGACCGCGGTGTTCATCGCGAACTGGCGCTCGCC 3191  
Qy 485 ProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAsp 504  
Db 3192 GCCCGCAGCGCAC-----GACTGCGCCACGCTCGAACAGCTCGAC 3233  
Qy 505 ValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal 524  
Db 3234 GTCACCTCCGTGCGCGCGATCCCGCGCGCAGCGGC---ACCGCGCAGACCTGGGTG 3290  
Qy 525 AspTyrAlaGluLeuAsnProSerIleValAlaGlySerPheAspProSerSer 544  
Db 3291 GAT-----GAAACCGCGCC 3305  
Qy 545 GlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly 564  
Db 3306 GACGGCGCGCGCTTCACCGTCCACCGCGTCCGCGACCGCGTGGAGCTGCAC 3365  
Qy 565 SerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg 584  
Db 3366 GCCGAG-----GGGTTCCTCCGCCCGCGCGTG---CCCCAGCCCGAAGCGTCGAC 3416  
Qy 585 PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn 604  
Db 3417 ACCGCTGGCGCGC-----CCGGC-----GGGTGCGCGCGACGGGTGCCCGGG 3464  
Qy 605 SerTrpAlaAlaSerGlnGlyValProAlaAlaGlnIleArgSerAspArgValAsn 624  
Db 3465 GCGTGGCGACCGCGGACCGAGTCTTCGTGAAGCCGAGTGCAGACG----- 3512  
Qy 625 ProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAsp 641  
Db 3513 CCTGACGGTTCGTGGCACACCCCGACCTGCTCGACGGGTCTTCTCCGCGTCCGCGAC 3572  
Qy 642 GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal 661  
Db 3573 GGG-----AGCCGCGACCGACCGGATG-GCGCGACCTCGCGGTGCACGCTCGACGC 3625  
Qy 662 MetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu 681  
Db 3626 CACCGTGTGCGCGCTCCCTCACCCCGCGACAGTGTGTGTCGAGCTCGCCGCTT 3685  
Qy 682 TyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAla--- 700  
Db 3686 CGACGGTCCGGATGCGCGT-----GCTACCGCGAGTTCGTGACGCTGGG 3733  
Qy 701 ---ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal 719  
Db 3734 CGAGTCCGCTCGCGCGCGATCCGACGAGTCGACGCTGCTTCGCTTGGAGTGT 3793  
Qy 720 ValGlyThrIleGlyGlyVal-ThrGlyAlaTyrArgSerAspCysGlyThrThrTr 739  
Db 3794 GCGGTGCGCGAGGCCACTACGAGCTGCC-----GACGAGCTCCCGAGGCGTA 3844  
Qy 739 pValIleAsn-----AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAl 755  
Db 3845 CACCTTCATCACCGCCACACACCCCGACCGACCCCGACCCACCCACACAC 3904  
Qy 755 ailerThrGlyAspHisAlaAsnLeuArgVal-----TyrIle 768  
Db 3905 ACCCACGCGACCCACACACAAACACAGCGTCTCTACCGCCCTCCACACACCTCAT 3964  
Qy 768 eGlyThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaProSe 785  
Db 3965 CACCACCAACCAACCTCATCTGTCTCCACACACACCGACCCCGCGCGCGCTGCAC 4024



Db 2979 ACCGATGGCGGCGCAACCCGATCGGCCAGGCCACCGCGATGGCAGCGCAACTGGAGC 2920  
Qy 391 SerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla----- 405  
Db 2919 TTCACCGCGGCGACCGCGTGGCGCAACCGCACCGGTGATCAATCGGTGCCCGCAGGATCCG 2860  
Qy 406 -----GluProTrpLeuThrPhe-----GlyValGlnProAsnProPro 418  
Db 2859 GCCGGCAATACACGCGGCCCGACAGCACCGACCGGTGGACCGGTGGCGGCCCGCCACCCCG 2800  
Qy 419 ValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSer 438  
Db 2799 GTGGTCAACCGCACCGCC----- 2779  
Qy 439 AspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAspLeu 455  
Db 2778 ---AGCGTGTACCGCGGTACCGCGGAAGCGCGGCCACCGGTGATCTCACC----- 2731  
Qy 456 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 475  
Db 2730 -----GACGGCGCGCGCAAC-----CCGATCGCGCAGGTCAACCGCCGAC 2692  
Qy 476 ThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu----- 490  
Db 2691 GGCAGCGCAACTGGAGCTTCACGCC-----GGCACCGCGCTGGCCAAACGGCTCGGTG 2638  
Qy 491 IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAlaVal 508  
Db 2637 ATCAATGGCTGGCCAGGACCGCGCGCAACACCGCGGCCCGCCAGCACCGGTG 2578  
Qy 509 ProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerValAsp--- 525  
Db 2577 GACTCGGTAGTCCCGCGCACCGCGTGTCTGATCCGAGCAACCGGTACGGTATCAGCGGT 2518  
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545  
Db 2517 ACCCGCAAGCGGGCGCGACCGGTATCTCCACGACGCGCGCGCAACCCGATACGGCAG 2458  
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565  
Db 2457 GCC-----ACCGCGATGGCAGCGCAACTGG-----AGCTTC 2425  
Qy 566 GluProGlyGlyValThrGlyGlyVal-----AlaAlaSerAlaAspGlySer 583  
Db 2424 ACTCGGGCACCGCTGACCAACCGCACCGGTGATCAATCGGTGGCCCGCCAGCACCGCC 2365  
Qy 584 ArgPheValTrpAlaPro----- 589  
Db 2364 GGCACACACGCGGTCCGGTCAGCACACAGTGGACGCGGTGGCCCGCCACCGCGGTG 2305  
Qy 590 -----GlyAspProGlyGlnProValVal 597  
Db 2304 ATCGACCGGAGCAATGTGTCAAACTCAGCGCACCGCGCAACCGCGGTCCGGGTGATC 2245  
Qy 598 TyrAlaValGlyPheGlyAsn-----SerTrp 606  
Db 2244 CTCACCGATGGCAATGGCAACCCGATCGGCCAGACCCCTCGCCGACGAGTAGGCGCACTGG 2185  
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 626  
Db 2184 ACCTTCACCGCGCACCGCG----- 2164  
Qy 627 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 646  
Db 2163 -----CTGGCCCAACGGCACCGGTGTCTCAACGCC-----GTGGCCCGAC 2128  
Qy 647 GlnProValAla---AlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla 665  
Db 2127 GACCGCGCGGCAATACACGCGGTCCGGCCAGCACCGGTGGATACGTTGGCGCGCGCC 2068  
Qy 666 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThr 685  
Db 2067 ACGCG-----GTGATCAATGCCAGC 2047

Qy 686 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPhe 705  
Db 2046 AACGGC-----AGCGTGTATCACCGGACCGCC-----GAGGTGCGC--- 2011  
Qy 706 GlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGly 725  
Db 2010 -----GCCAAGATGATCTCACCGCAGCGCAACGGCAAC 1978  
Qy 726 ValThrGlyAlaTyrArgSerAspAspGlyThrThrTrpValLeuIleAsnAspAsp 745  
Db 1977 CCGATCGCGGACACACCGCGACCGCATGGC---AATGGACCTTCACCCCGGACG 1921  
Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765  
Db 1920 CCGCTGGCCAAACGCTGATCAACGCGGTCCGCGCAAGACCGCGC----- 1873  
Qy 766 ValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaProSer 785  
Db 1872 -----GGCAACGCCACGCGT---CCGGCC 1852  
Qy 786 GlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerProSe 805  
Db 1851 AGCACACGCTGACTCGGTGGCGCGTCCGCT---CCGCTGTGACATCAGCGCGCAC 1795  
Qy 805 rProSerSerProSerPro-----SerProSerProSe 817  
Db 1794 GCGCGCTGCTGACCGGACCGCCGAGCGACAGTCAGGTGCGTATCGTGGTCAACGGC 1735  
Qy 817 rSerSerProSerSerProSerProSerProSerProSerProSerProSerArgSe 837  
Db 1734 GACACCGCCAAACCGATCACGCTCACCGTCGACGCGCGCGCAACTTCAGCTCGCGTTC 1675  
Qy 837 r-----ProSerPr 840  
Db 1674 GCGCCGCGCTGATCACCGCGCAACTGATCGCGGGTTCCTCGTGGACCGCGCGCAAC 1615  
Qy 840 sSerAlaSerProSerProSerProSerProSerProSerProSerProSer---Se 859  
Db 1614 GTCACGCGCGCGCCACCATCAACGCGCGGACCTGGCGCGCGGACCATCAGGTGCCG 1555  
Qy 859 rProSerProThrProSerSerProValSerGlyValValValGlnTyrLysAs 879  
Db 1554 GAAGCGCGGATACCTGGATCAACCGCGGAGATCGGGGACGGC-ATCCAGGTC----- 1501  
Qy 879 rAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrG 899  
Db 1500 -----GATGTGACGCTCGCTCCGACCATCAGGTGCC----- 1468  
Qy 899 ySerSerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyG 919  
Db 1467 -----CAGGTGTACCGTCAAGTTTC---GCCGGGCAAGACGGCTA 1430  
Qy 919 ySerSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAl 939  
Db 1429 CGAGCGCGGAGTACCCATACCTCACCGCGGCGATCGCGCGGCAACCTGACCT 1370  
Qy 939 sSerPhe-----GlySerValAsnPro-----AlaThrProThrAl 951  
Db 1369 GACCTGACGCTCCCGCGCGCATGGCGCGCTTCGCGAGGGCGCGCTCGACCGCTCACCGC 1310  
Qy 951 aAsp 952  
Db 1309 CGAC 1306

Search completed: March 2, 2006, 12:21:12  
Job time : 512.215 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:19:26 ; Search time 1796.03 Seconds  
(without alignments)  
4406.259 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5134

Sequence: 1 MDRSENIRLWRSRLVSL.....RASFGSVNPATPTADTVLQX 957

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlp  
-Q-/abs/ABSSWEB.spool/US09917376/runat\_02032006\_091500\_8223/app\_query.fasta\_1  
-DB=Published Applications NA Main -QWMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTENT=pct -NORM=ext  
-HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs06p  
-USER=US09917376 @CGN 1 1 3095 @runat\_02032006\_091500\_8223 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score  | Query Match | Length  | ID | Description        |
|------------|--------|-------------|---------|----|--------------------|
| 1          | 5134   | 100.0       | 2869    | 3  | US-09-917-376-2    |
| 2          | 5134   | 100.0       | 2869    | 5  | US-10-155-400-2    |
| 3          | 2487   | 48.4        | 9025608 | 6  | US-10-156-761-1    |
| 4          | 2478   | 48.3        | 2646    | 6  | US-10-156-761-1845 |
| 5          | 1631.5 | 31.8        | 2517    | 5  | US-10-026-994-4    |
| 6          | 1631.5 | 31.8        | 2710    | 5  | US-10-026-994-1    |
| 7          | 1570   | 30.6        | 9025608 | 6  | US-10-156-761-1    |

Alignment Scores:  
Pred. No.: 0  
Score: 5134.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Length: 2869  
Matches: 956  
Conservative: 0  
Mismatches: 0

#### ALIGNMENTS

RESULT 1

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOLYTIC

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2869

; TYPE: DNA

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; NAME/KEY: modified\_base

; LOCATION: (2869)

; OTHER INFORMATION: a, c, t, g, other or unknown

; US-09-917-376-2

8 1442 28.1 2217 6 US-10-156-761-2561 Sequence 2561, Ap  
9 1167.5 22.7 3668 3 US-09-927-827-21 Sequence 21, Appl  
10 1125.5 21.9 2646 7 US-10-395-241-11 Sequence 11, Appl  
11 1124 21.9 2481 7 US-10-395-241-17 Sequence 17, Appl  
12 1117 21.8 2367 7 US-10-395-241-13 Sequence 13, Appl  
13 940 18.3 5698 7 US-10-420-191-1 Sequence 1, Appli  
14 823 16.0 3687 3 US-09-917-384-2 Sequence 2, Appli  
15 823 16.0 3687 3 US-09-917-383-2 Sequence 2, Appli  
16 702 13.7 2289 8 US-09-917-378-2 Sequence 2, Appli  
17 392.5 7.6 1103 8 US-10-653-047-7511 Sequence 7511, Ap  
18 365.5 7.1 651 9 US-10-450-763-20760 Sequence 20760, A  
19 361 7.0 2223 6 US-10-156-761-550 Sequence 550, App  
20 346.5 6.7 2706 9 US-10-939-262-11 Sequence 11, Appl  
21 345 6.7 18596 3 US-09-880-107-1590 Sequence 1590, Ap  
22 345 6.7 18596 3 US-09-967-768A-119 Sequence 119, App  
23 345 6.7 18596 3 US-09-954-531-124 Sequence 124, App  
24 345 6.7 18596 3 US-09-954-531-348 Sequence 348, App  
25 345 6.7 18596 8 US-10-629-313-140 Sequence 140, App  
26 345 6.7 18596 9 US-10-843-641A-1191 Sequence 1191, Ap  
27 345 6.7 18596 9 US-10-843-641A-1415 Sequence 1415, Ap  
28 345 6.7 18596 9 US-10-843-641A-6264 Sequence 6264, Ap  
29 331 6.4 1887 9 US-10-939-262-13 Sequence 13, Appl  
30 330.5 6.4 7407 6 US-10-246-330-3 Sequence 3, Appli  
31 330.5 6.4 7407 7 US-10-282-122A-30151 Sequence 30151, A  
32 319.5 6.2 9495 8 US-10-669-161-71 Sequence 71, Appl  
33 299.5 5.8 210528 5 US-10-087-192-289 Sequence 289, App  
34 297 5.8 32591 6 US-10-085-117-187 Sequence 187, App  
35 296.5 5.8 4767 3 US-09-940-316B-28 Sequence 28, Appl  
36 296 5.8 96602 6 US-10-085-117-61 Sequence 61, Appl  
37 293 5.7 4818 3 US-09-940-316B-32 Sequence 32, Appl  
38 292.5 5.7 185548 5 US-10-175-523-62 Sequence 62, Appl  
39 292.5 5.7 185548 10 US-11-099-266-62 Sequence 17, Appl  
40 290 5.6 1232 7 US-10-761-169-17 Sequence 17, Appl  
41 289 5.6 324 7 US-10-282-122A-13296 Sequence 13296, A  
42 288.5 5.6 4571 3 US-09-940-316B-18 Sequence 18, Appl  
43 288 5.6 288 6 US-10-437-708-111 Sequence 111, App  
44 288 5.6 288 6 US-10-395-402-111 Sequence 111, App  
45 288 5.6 288 8 US-10-257-199-111 Sequence 111, App



|  |        |  |      |
|--|--------|--|------|
| Query Match:                                       | 100.0% | Indels:  | 0    |
| DB:  | 3      | Gaps:  | 0    |
| US-09-917-376-1 (1-957) x US-09-917-376-2 (1-2869) |        |  |      |
| QY   | 1      | MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu     | 20   |
| DB   | 1      | ATGATCGTTTCGGAGAACATCCGCTGACTATGATGATCAGACAGATGGTATCATCTGCTC     | 60   |
| QY   | 21     | AlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThr        | 40   |
| DB   | 61     | GCGCCACTGCTGCTTGCCTGCGCGTGGCCGCTCTGGGAGTTCTGCCCATCGGATAACG       | 120  |
| QY   | 41     | AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly     | 60   |
| DB   | 121    | GCCTTCTCTCGCGCACCGCGGCAGCACTCAGCGGTACACCTGGAGCAACGTGGCATCGG      | 180  |
| QY   | 61     | GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal     | 80   |
| DB   | 181    | GCCTGCGGCTTTGTGACGCGGATCGTCTTCAATGAAGGTGCACCGGGAATTCCTGTACG      | 240  |
| QY   | 81     | ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu     | 100  |
| DB   | 241    | CGGACGACATCGGGGATGTATCGATGGATGCGGCAACGGCGGTGATCCCTCTT            | 300  |
| QY   | 101    | LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAsp     | 120  |
| DB   | 301    | CTGGATTGGTGGATGGAACAAATTGGGGTACAAACGCGCTCGTCAGCATTCGCGCAGAC      | 360  |
| QY   | 121    | ProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPro     | 140  |
| DB   | 361    | CCGATCAATACTAAACAGGTATGGCGCGCGCTCGGAATGTACACCAACAGCTGGGACCCA     | 420  |
| QY   | 141    | AsnAspGlyAlaIleLeuAspSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeu     | 160  |
| DB   | 421    | AACGACGAGGATTCCTCGCTCTGTATCAGGGCGCAACGTGGCAATAACGCCCTG           | 480  |
| QY   | 161    | ProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp     | 180  |
| DB   | 481    | CCGTTCAAGCTTGGCGCAACATCCCGGGCGTGGATGGCGAGCGCTTCGCGTGGAT          | 540  |
| QY   | 181    | ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSer     | 200  |
| DB   | 541    | CCAAACAATGACAAATCTGTATTCGGCGCCCGAGCGCAAGGGCTCTCGAGAAGC           | 600  |
| QY   | 201    | ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle     | 220  |
| DB   | 601    | ACGATTCGCGCGGACCTGCTCCAGATGACGAACCTTTCGGACGTAGGCACCGTACAT        | 660  |
| QY   | 221    | AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla     | 240  |
| DB   | 661    | GCAATCCCACTGACACGCGGTATCAGCGCATATTCAGGGCTGCTCTGGGTGCT            | 720  |
| QY   | 241    | PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla        | 260  |
| DB   | 721    | TTGCACAAGTCTTCGTATCGCTCGGCAACGAGTAAGACCAATTTTGTGGCGTGGCG         | 780  |
| QY   | 261    | AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro     | 280  |
| DB   | 781    | GATCCCAATAATCCGGTCTTCGTGAGCAGAGACGGCGCGGACGTGGCAGCGGTGCGG        | 840  |
| QY   | 281    | GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu     | 300  |
| DB   | 841    | GGTGGCGGACCGGCTTCATCCCGCAAGGGCGTCTTTGACCGGTCAACACGCTGCTC         | 900  |
| QY   | 301    | TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys     | 320  |
| DB   | 901    | TATATTGCCACCAATACGGTGGTCCGTATCAGCGGAGCTCGGCGAGCTCTGAAA           | 960  |
| QY   | 321    | PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla     | 340  |
| DB   | 961    | TTCTCGGTGACCTCCGGGACATGACGCGAATCAGCCCGGTACCTTCGACGACACGCGC       | 1020 |
| QY   | 341    | AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet     | 360  |
| DB   | 1021   | AACGACTACTTTTGGTTACAGCGCGCTCACTATCGACCGCGCAGCACCCGCAACGATAATG    | 1080 |
| QY   | 361    | ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly     | 380  |
| DB   | 1081   | GTGGCAACCCAGATATCGTGGTGGCGGACACCAATAATTTTCGGAGCACCGACGGGT        | 1140 |
| QY   | 381    | AlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrVal     | 400  |
| DB   | 1141   | CGGACGTGGACCGGATCTGGGATTTGGAGTTGACGAGTTATCCCAATCGAAGCTTCGATATGTG | 1200 |
| QY   | 401    | LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro     | 420  |
| DB   | 1201   | CTTGACATTTGGCGGAGCTTGGCTGACCTTCGGCGTACAGCGGAATCTCTCCGTATCCC      | 1260 |
| QY   | 421    | SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg     | 440  |
| DB   | 1261   | AGTCCGAAGCTCGGCTGGATGGATGAAGCATGGCAATCGATCCGTTCAACTCTGTATCGG     | 1320 |
| QY   | 441    | MetLeuTyrGlyThrGlyValaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer    | 460  |
| DB   | 1321   | ATGCTCTACGACACAGCGCGGACGTTGTACGCAACAATGATCTCAGAGTGGGACTCC        | 1380 |
| QY   | 461    | GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp        | 480  |
| DB   | 1381   | GGCGGCCAGATTCAATATCGCGCGATGTCAAAGGATTTGGAGGAGACGCGGTAAACGAT      | 1440 |
| QY   | 481    | LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPhe        | 500  |
| DB   | 1441   | CTCATCAGCCCGCGCTCTGGCGCGCTCATCAGCGCTCTCGGACCTCTCGCGCGCTTC        | 1500 |
| QY   | 501    | ThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThr     | 520  |
| DB   | 1501   | ACCCACGCGAGCTTACTGCGGTGCTCATCGAGATCTTCACGTACCGGTGTTACGACC        | 1560 |
| QY   | 521    | GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe     | 540  |
| DB   | 1561   | GGCACCGGTGCGATATGCGGAATTTGAATCCGTGATCATCTGTCGCGCTGGAAGTTTC       | 1620 |
| QY   | 541    | AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn     | 560  |
| DB   | 1621   | GATCCATCGAGCCAAACCGAACGACAGGACGCTCGCGTCTTCGACAGACGCGCGCAAGAAC    | 1680 |
| QY   | 561    | TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla        | 580  |
| DB   | 1681   | TGTTTCCAAAGCAGCGAACCTTGGCGGGGTGACGACGGCGCGCACCGTCCGCCATCGGC      | 1740 |
| QY   | 581    | AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTrpAlaVal     | 600  |
| DB   | 1741   | GACGCTCTCGTTTCTGCTGCGGTCCCGCGATCCCGGTACGCTGTGTGTGTACGCGATC       | 1800 |
| QY   | 601    | GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer     | 620  |
| DB   | 1801   | GGATTTGGCAACTCTCTGGGTGCTTCGCAAGGTGTTCCGCCAATGCCCAGATCCGCTCA      | 1860 |
| QY   | 621    | AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr     | 640  |
| DB   | 1861   | GACCGGTGAATCCAAAGACTTTCTATGCCCTATCCAATGGAACCTTCTATCGAAGCAGG      | 1920 |
| QY   | 641    | AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly     | 660  |
| DB   | 1921   | GACGCGCGGTGACATTCCAACCGGTGCGCGCGGTCTTCGAGCAGCGGTGCCGTGCT         | 1980 |
| QY   | 661    | ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly     | 680  |
| DB   | 1981   | GTGATGTTCCACGCGGTGCTTGGAAAAGAGCGGATCTGTGGCTGCTGCTGATCGAGCGG      | 2040 |
| QY   | 681    | LeuTyrHisSerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSerAla        | 700  |
| DB   | 2041   | CTTTACCACTCAACCAATGGCGGACGAGTTGGTCTGCAATCACCAGCGGTATCTCCGCG      | 2100 |
| QY   | 701    | ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal     | 720  |



```
Db 2101 GTGACGTGGGATTGGTAACTCGCGCGGGTCTCATACCGACCGCTTTGTCTGCTC 2160
Qy GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpVal 740
Db 2161 GGCACGATCGGAGCGTTACGGGGCGGTACCGCTCCGACGACTGTGGACGACCTGGGTA 2220
Qy LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760
Db 2221 CTGATCAATGATGACCAACCAATACGGAATTTGGGACAAAGCAATCACCCTGGTACAC 2280
Qy AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 2281 GCGAATTTACGGCGGGTGTACATAGCAGCAAGCGCGGTGGAAATGTATACGGGACAAAT 2340
Qy GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800
Db 2341 GGTGTGCGCGCTCGGATCGCGTCTCGTCCGTGAGTCCGTCCGCTTCGCCGAGCCTG 2400
Qy SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820
Db 2401 AGCCGAGCCGAGCCGAGCAGCTCGGCATCGCGCTCGCGCTCGCGAGCTCGAGTCCA 2460
Qy SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840
Db 2461 TCCTCGTCCGCTCTCCGCTCGCGCTCACCATCGCGAGTCCGCTCTCCGTCACCA 2520
Qy SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860
Db 2521 TCGGGCTCGCGAGCCCGTCTTCGTACCGAGCCCGTCTTCGTACCCGCTCTTCGTCGCG 2580
Qy SerProThrProSerSerSerProValSerGlyGlyValIleValIleValIleValIle 880
Db 2581 AGCCCAACGCGCTCGTCTCGCGCGTGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 2640
Qy AspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySer 900
Db 2641 GATTCCGCGCGCGGTGTAATACAGTCAAGCGCGGTTCAGGTGCTGTAATACCGGGTGC 2700
Qy SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920
Db 2701 TCGTCCGTTGGATTGTCCAGCGGTGACGTGCGGTACTCGTTTACCCGCGGATGGTGC 2760
Qy SerThrLeuValTyrAsnCysAspTrpAlaIleGlyCysGlyAsnIleArgAlaSer 940
Db 2761 TCGACACTGGGTGACAACTGTGACTGGCGGCGGATCGGGTGTGGGAATATCCGCGCTCG 2820
Qy PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 2821 TTGGCTCGGTGAACCCGCGAGCCGCGAGCGCGGACACTTACCTGCAG 2868
```

## RESULT 2

```
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU.
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
```

```
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2
```

## Alignment Scores:

```
Pred. No.: 0 Length: 2869
Score: 5134.00 Matches: 956
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 5 Gaps: 0
```

US-09-917-376-1 (1-957) x US-10-155-400-2 (1-2869)

```
Qy 1 MetAspArgSerGluAsnIleArgLeuThrMetArgSerArgArgLeuValSerLeuLeu 20
Db 1 ATGGATCGTTCCGAGAACATCGCTCTGACTATGAGATCAGCAGATTGGTATCACTGCTC 60
Qy 21 AlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuProIleAlaIleThr 40
Db 61 GCGCGCACTGCGTCTGTCGCGTGGCGCGCTCTCGGAGTTCTGCGCCATCGCGATAACG 120
Qy 41 AlaSerProAlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGly 60
Db 121 GCTTCTCTCGGCACGCGCGCAGACTCAGCCGTACACCTGGAGCAACGTGGCGATCGGG 180
Qy 61 GlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrVal 80
Db 181 GCGCGCGCTTGTTCGACGGGATCGTCTTCATGAAGTGCACCGGGAATTCGTACGTG 240
Qy 81 ArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeu 100
Db 241 CGGACGGACATCGGGGGATGTATCGATGGGATCGCCCAACGCGCGGTGGATCCCTCTT 300
Qy 101 LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValIleSerIleAlaAla 120
Db 301 CTGGATTGGGTGGATGGAACAAATTTGGGGGTACAACGCGCTGTCTACAGCATTTGCGG 360
Qy 121 ProfileAsnThrAsnLysValTrpAlaValGlyMetTyrThrAsnSerTrpAspPro 140
Db 361 CGGATCAATACTAAACAGGTATGGCGCGCTCGGGAATGTACACCAACAGCTGGGACCCA 420
Qy 141 AsnAspGlyAlaIleLeuArgSerSerAspGlnGlyValAlaThrTrpGlnIleThrProLeu 160
Db 421 AACGACGGAGCGATTCTCCGCTCGTCTGATCAGGCGCAACGTGGCAATAAACGCCCTG 480
Qy 161 ProphelLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyuArgLeuAlaValAsp 180
Db 481 CCGTTCAAAGCTTGGCGCAACATGCGCGCGCTGGAAATGGCGAGCGGCTTGGCGTGGAT 540
Qy 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyGlyGlyLeuTrpArgSer 200
Db 541 CCAAAACAATGACAACATTCTGTATTTCGGCGCGCCCGCGGCAAGGGCTCTCGGAAGC 600
Qy 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220
Db 601 ACAGATTCCGGCGGACCTGTGTCAGATGACGAACCTTTCCGAGCTGAGCAGCATATT 660
Qy 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAla 240
Db 661 GCAATCCCACTGACACGACCGGCTATCAGAGCGATATTCAAGGCGTCTGCTGGGTCGCT 720
Qy 241 PheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260
Db 721 TTCGACAAAGTCTTCGTCTCGTCCGGCAAGCGAGTAAGACCAATTTTGTGGCGGTGGCG 780
Qy 261 AspProAsnAsnProValPheTrpSerArgAspGlyValAlaThrTrpGlnAlaValPro 280
Db 781 GATCCCAATATCCCGTCTTCTGGAGCAGAGACGCGCGCGCAGCGGCGGCGGCGGCGG 840
Qy 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300
Db 841 GGTGGCGCGACCGGCTTTCATCCCGCACAGGGCGCTTTTGACCCCGGTCAACACGCTGCTC 900
```

Qy 301 TyrileAlaThrSerAsnThrGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
Db 901 TATATTGCCACACCAATACCGGTGGTCCGATACCGGAGCTCCGGCAGCTCGGAAA 960  
Qy 321 PheSerValThrSerGlyThrTrpThrArgileSerProValProSerThrAspThrAla 340  
Db 961 TTCTCGGTGACTCCGGACATGACCGGAATACGCCGGTACCTTCGACGGACACGGCC 1020  
Qy 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
Db 1021 AACGACTACTTTGGTTACAGCGGCTCCTATACGCCCGCAGCACCGACACCATATG 1080  
Qy 361 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
Db 1081 GTGCAACCCAGATATCGTGTGGCCGACACCAATAATCTTCGGAGCACCGCGCGGT 1140  
Qy 381 AlaThrTrpThrArgileTyrAspThrTrpSerTyrProAsnArgSerLeuArgTyrVal 400  
Db 1141 GCGACGTGGACCGGATCTGGGATGGACGAGTTATCCCAATCGAAGCTTCGCGATATGTG 1200  
Qy 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProValPro 420  
Db 1201 CTTGACATTTCCGGGAGCCTTGGCTGACCTTCGCGGTACAGCGAATCTCCCGTACCC 1260  
Qy 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
Db 1261 AGTCCGAAGCTCGGCTGGATGAAGCGATGGCAATCGATCCGTTCAACTCTGATCGG 1320  
Qy 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
Db 1321 ATGCTCTACGGAACAGCGCGGACGTTGTACGCAAAATATCTCACGAAGTGGGACTCC 1380  
Qy 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480  
Db 1381 GCGCGGCAGATTCTATATCGCGCCGATGTTAAAGATTGGAGGAGCGCGGTAAACGAT 1440  
Qy 481 LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe 500  
Db 1441 CTCATCAGCCCGCGTCTGGCGCCCGCTCATCAGCGCTCTCGGAGACCTCCGGCGGCTC 1500  
Qy 501 ThrHisAlaSerValThrAlaValProSerThrIlePheThrSerProValPheThrThr 520  
Db 1501 ACCCAGCCCGCAGTTACTGCGGTGCCATCAGCAGATCTTCAGTCACCGGTGTTCCAGC 1560  
Qy 521 GlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPhe 540  
Db 1561 GGCACCGCGTGGACTATGCGGAATTGAATCCGTCGATCATCGTTCGCGTGGAGTTTC 1620  
Qy 541 AspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsn 560  
Db 1621 GATCCATCGAGCCCAACCGACGACGACGCTCGCGTCTTCGACAGACGGCGGCAAGAC 1680  
Qy 561 TrpPheGlnGlySerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAla 580  
Db 1681 TGGTTCCAAAGCAGCAACCTGGCGGGGTGACAGCGGCGGCACCGTCGCGCATCGGCC 1740  
Qy 581 AspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 600  
Db 1741 GACGGCTCTCGTTTCGTTCCGCTCCCGCGCATCCCGGTGAGCTGTCGCGTGTACGCGAGTC 1800  
Qy 601 GlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSer 620  
Db 1801 GGATTTGGCAACTCTCGGTGCTTCCCAAGGTGTTCCCGCAATGCCAGATCCCGCTCA 1860  
Qy 621 AspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThr 640  
Db 1861 GACCGGTGATCCAAAGACTTCTATGCGCTATCCATGGAACCTTCTATCGAAGACG 1920  
Qy 641 AspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGly 660  
Db 1921 GACCGCGCGTGCATTTCCAAACCGGTTCGCGCGGTCTTCGAGCAGCGCGTCCGCTCGGT 1980

Qy 661 ValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGly 680  
Db 1981 GTCATGTTTCCACGCGGTGCTCGAAAAGAGCGCATCTGTGGCTCGTCATCGAGCGGG 2040  
Qy 681 LeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerAla 700  
Db 2041 CTTTACCACACTCAACATATGGCGGACAGTTGGTCTGCAATCACCGGGGTATCTCCCGG 2100  
Qy 701 ValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 720  
Db 2101 GTGAACGTGGGATTTGGTAAGTCTCGGCCCGGTCGTATACCCAGCGCTCTTGTCTGTC 2160  
Qy 721 GlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpVal 740  
Db 2161 GGCACGATCGGAGCGTTTACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTA 2220  
Qy 741 LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHis 760  
Db 2221 CTGATCAATGATGACCAAGCAATACGAAATTTGGGACCAAGCAATCACCGGTGACCAAC 2280  
Qy 761 AlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780  
Db 2281 GCGAATTTACGCGGGTGTATACAGGACCAAGCGCGTGGAAATGTATACGGGACATT 2340  
Qy 781 GlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSerLeu 800  
Db 2341 GGTGGTCCGCGTCCGGATCCGCTCCGTCGCTGAGTCCGCTTCGCGAGCCTG 2400  
Qy 801 SerProSerProSerProSerProSerProSerProSerProSerProSerProSerPro 820  
Db 2401 AGCCGAGCCGAGCCCGAGCAGCTCGCCATTCGCGCTCGCGCTTCGCGAGCCTG 2460  
Qy 821 SerSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 840  
Db 2461 TCCTGTCGCGCTCTCCGTCGCCGTCACCATCGCGAGTCCGTCCTCGGTCTCCGTCACCA 2520  
Qy 841 SerAlaSerProSerProSerProSerProSerProSerProSerProSerProSerPro 860  
Db 2521 TCGGCGTCCGCGAGCCGCTTCGTCACCGAGCCGCTTCGTCACCGCTCTTCGTCGCGG 2580  
Qy 861 SerProThrProSerSerSerProValSerGlyGlyValLysValGlnTyrLysAsnAsn 880  
Db 2581 AGCCCAACGCGCTCGTCCGCGGTGTGGGTGGGTGGAAGGTGTCAGTATAAGAAATAAT 2640  
Qy 881 AspSerAlaProGlyAspAsnGlnIleIleProGlyLeuGlnValValAsnThrGlySer 900  
Db 2641 GATTCCGGCGCGGTGATAATCAGATCAAGCCGGGTTTTCAGGTTGGTAATACCGGGTGG 2700  
Qy 901 SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGlyGlySer 920  
Db 2701 TCGTCGTTGGATTTGTGACCGGTGACGGTACGCTGCTGTTTACCCCGGATGGTGGCTCG 2760  
Qy 921 SerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSer 940  
Db 2761 TCGACACTGGTGTACAACTGTGACTGGGCGGCGATCGGCTGGGAAATATCCGCGCTCG 2820  
Qy 941 PheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 2821 TTCGGCTCGGTGAACCCCGCGACGCCGCGGACACCTACCTGTCAG 2868

## RESULT 3

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

|    |         |  |         |
|----|---------|--|---------|
| Db | 2278033 | AGCGGGTACGCTCCGACAAACACGAGGCGATCGTGTGGTCACCTTCACGAGTTCGACGGG | 2278092 |
| Qy | 247     | SerLeuGlyGlnAlaSerIysThrIlePheValGlyValAlaAspProAsnAsnProVal | 266     |
| Db | 2278093 | TCGCGGGGAGCTCCACGCGGACGGTGTAGCTCGGGGTCCGCCAAGACAACTCCGTC     | 2278152 |
| Qy | 267     | PheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe    | 286     |
| Db | 2278153 | TATCGCTCCACGACGCGGGCGGACCTGGTCCGGCTGGCGCCAGCCACCGGCCAT       | 2278212 |
| Qy | 287     | IleProHisIysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn | 306     |
| Db | 2278213 | CTCGCCACAAAGGGCGTGTGGACGCGCGCAACCGCTGTCTGTACTCGCGTACACGCAC   | 2278272 |
| Qy | 307     | ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpIysPheSerValThrSerGly | 326     |
| Db | 2278273 | AAGGGCGGACCGTACGACGCGCGCAAGGACACCTGTGGCGGTACAGCAAGACGCGG     | 2278332 |
| Qy | 327     | ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr | 346     |
| Db | 2278333 | ACCTGGACGAACATCAGCCGGTCGCGAGGCGCACACC-----TACTACGGCTTC       | 2278383 |
| Qy | 347     | SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer | 366     |
| Db | 2278384 | AGCGGGCTGACCGTGGACCGGACGATCCGGGACCGTGATGGCGACTGCGTACAGCTCC   | 2278443 |
| Qy | 367     | TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle | 386     |
| Db | 2278444 | TGTGGCCGACACGCGAGCTCTTCCGCTCCACGGACAGCGCGGCACTTGGACGAAGGCC   | 2278503 |
| Qy | 387     | TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu | 406     |
| Db | 2278504 | TGGGACTACACCTCGTATCCGAGCCGCTGAACCGCTTACCATGGATGCTCTGCTCTCG   | 2278563 |
| Qy | 407     | ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProIysLeuGlyTyr | 426     |
| Db | 2278564 | CCCTGGCTCACCTGGGGAGCGCAACCCCGCACCGCCGAGCAGACCCCGAACTCGGCTCG  | 2278623 |
| Qy | 427     | MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly | 446     |
| Db | 2278624 | ATACCGAGTCCCTGGAGATCGACCGCTTCGACTCCGCGCGCAATGATACGGAAACGGGC  | 2278683 |
| Qy | 447     | AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle | 466     |
| Db | 2278684 | GCACGGTCTACGGCACGGACCACTGACGACTGGGACAGCGGAAGCCAGTTCACCATC    | 2278743 |
| Qy | 467     | AlaProMetValIysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSer | 486     |
| Db | 2278744 | AAGCCGATGGCGGGCGCTTGGAGGAGACGGCCGTCAACGACCTCGCTCGCTCCCTCC    | 2278803 |
| Qy | 487     | ---GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal | 505     |
| Db | 2278804 | GGCGGGCGCCAGCTGTTCAGCGCGCTCGGTGACATCGCGGGCTTCGGGCACACGGACCTC | 2278863 |
| Qy | 506     | ThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerValAsp    | 525     |
| Db | 2278864 | ACCACGGTCCGCTCGCTGATACAGCTGCCGAACCTTACCACAGACCACTCGCTCGAC    | 2278923 |
| Qy | 526     | TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln | 545     |
| Db | 2278924 | TACCGCGAGACCCCGGGCACGGTGGTGGCGGTTCGGCAATCTCGAC-----TCGGGT    | 2278977 |
| Qy | 546     | ProAsnAspArgHisValAlaPheSerThrAspGlyGlyIysAsnTrpPheGlnGlySer | 565     |
| Db | 2278978 | CCG-----CATGTGGCTTCTCGACGGCAACCGGGCGCAACTGGTTCGCGGGGGCGG     | 2279028 |
| Qy | 566     | GluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPhe | 585     |
| Db | 2279029 | GACCTTCGGGGGTACGGGGGTGGAGCGTTCGCGGCCGCGTCCACGCGCAGTCTCGCTTC  | 2279088 |
| Qy | 586     | ValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer | 605     |



QY 161 ProPheLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAsp 180  
DB 442 CCCTTCAAGCTGGCGGGAACAATGCCGGCCGGGCGATGGGGAGCGGCTCGCGTGCAC 501  
QY 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLyLeuGlyLeuTyrArgSer 200  
DB 502 CCGACAGAACAGCGGTCTGTATCTCGGGCGGCCACGGGCAAGGGCTGTGGCGGTGC 561  
QY 201 ThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle 220  
DB 562 ACGGACTCGGGGCTCTCTGTGTGCGAGGTCAACGACTTCCGAACTCGGCACCTACGTG 621  
QY 221 AlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAla 240  
DB 622 CAGGACGCGACGACACAGAGCGGTACCGTCCGACAAACAGGGGATCGTGTGGGTACC 681  
QY 241 PheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAla 260  
DB 682 TTCGACGAGTCCAGCGGTTCGCGGGAGCTCCACGGGACGGTGTACGTCTGGGTTCGC 741  
QY 261 AspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValPro 280  
DB 742 GACAAGGACACTCGTCTATCGCTCCACGAGCGGGCGGACCTGGTCCGGGTGGCC 801  
QY 281 GlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 300  
DB 802 GCGCAGCCACCGGCATCTCGCCCAAGGCGGTCTGGACCGCGGCAACGGCTGTCTG 861  
QY 301 TyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLys 320  
DB 862 TACCTCGGTACAGCAAGGGCGGACCGTACGACGGCGGCAAGGACAGCTGTGGCGG 921  
QY 321 PheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAla 340  
DB 922 TACACAGACACCGGACCTGGAGCAATACACCCGCTCCGAGGCGGACAC--- 978  
QY 341 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 360  
DB 979 -----TACTACGGCTTCAGCGGGCTGACCGTGGACCGGACATCCGGGACGGTGATG 1032  
QY 361 ValAlaThrGlnIleSerTrpProAspThrIleIlePheArgSerThrAspGlyGly 380  
DB 1033 GCGACTCGGTACAGCTCTCTGTGGCGGACACGACGCTCTTCGCTCCACGACGACGCGC 1092  
QY 381 AlaThrTrpThrArgIleTrpAspThrTrpSerTyrProAsnArgSerLeuArgTyrVal 400  
DB 1093 GGCACCTGGAGAGGCTGGGACTACACCTCGTATCCGAGCGCTCCGACCGCTCCACC 1152  
QY 401 LeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValPro 420  
DB 1153 ATGGATGTCTGTCTCGCCCTGGCTACCTGGGAGCGGAACCCGCGCCCGGAGCAG 1212  
QY 421 SerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 440  
DB 1213 ACCCGMAACTCGGTGTGATGACCGAGTCCCTGGAGATCGACCCGTTCGACTCCGCGCG 1272  
QY 441 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 460  
DB 1273 ATGATGTACGGAACGGGCGGCGACGCTACGCGACGAGACCTGACGAACTGGGACACG 1332  
QY 461 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp 480  
DB 1333 GGAAGCCAGTTCCATCAAGCCGATGGCGGGGCTGGAGGAGCGGCGCTCAACGAC 1392  
QY 481 LeuIleSerProProSer---GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 499  
DB 1393 CTCGCTCGCTCTCTCGCGGCGGCCAGCTGTTCAGCGCGCTCGGTGACATCGCGCGC 1452  
QY 500 PheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 519  
DB 1453 TTCGCGCACACGACCTCACACGCTGCGTGTGATGATACACGTCGCGCAACTTCACC 1512  
QY 520 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 539

DB 1513 ACAGACACCAAGCTCGACTACCGCGGAGACCGCCCGGCGACGCTGTGCGGGTCCGCAAT 1572  
QY 540 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 559  
DB 1573 CTCGAC-----TCGGGTCCG-----CATGTGGCGTTCGACGAGCAACGCGGCC 1617  
QY 560 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 579  
DB 1618 AACTGTGTTCCGGGGCGGACCTTCGGGGGTGAGCGGGGTGGACGCTCGGGCGCGC 1677  
QY 580 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 599  
DB 1678 TCCGACGCGAGTCGCTCTGTGTGGAGCCCGCGGCGCACCGGG-----GTGCAGTACACG 1731  
QY 600 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619  
DB 1732 ACCGGTTTCGGACCTCTGTGTGGGTCCGGGGCTCCCGCGGGGCGATCGTCGAG 1791  
QY 620 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 639  
DB 1792 TCCGACCGGGTCGACCCGAAGACCTTCTACGGCTTCAAGTCCGCGAGGTTCTACGTCA 1851  
QY 640 ThrAspGlyGlyValThrPheGlnProValAlaAla---GlyLeuProSerSerGlyAla 658  
DB 1852 TCGGACGGCGGGCGACCTTCACGGCTCCGGCGGCCACGGCGCTCGGACGCGCGACG 1911  
QY 659 ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer 678  
DB 1912 -----GTGCGCTTCAAGCGCTGCCCGGACGAGGGCGACATCTGGCTGGCGGGCGC 1965  
QY 679 Ser-----GlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla 693  
DB 1966 GCGAGCGACGGCGCTACGGGTGTGGCACTCGACGACGCGCGCGCGCTTCCACCAAG 2025  
QY 694 IleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 713  
DB 2026 CTCGCGACCGTCCGACCGGCGCACCATCGGCTTCGGCAAGCGCGACCGCGCTCGC 2085  
QY 714 TyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAsp 733  
DB 2086 TACCAAGACCTCTACACCGACGCGAAGATCGCGGTGTGGCGGCACTCTCCGCTCCACC 2145  
QY 734 AspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGly 753  
DB 2146 GACAGGGCGCGAGCTGGACCCGCTCAACGACGATGCCACCGATGGGGTTGACGCGGC 2205  
QY 754 GlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArg 773  
DB 2206 GCGGCGATCACCGGTGACCCCGGCTACGGGCGCGTGTATGTGTGACGACGAGCGGCGC 2265  
QY 774 GlyIleValTyrGlyAspIleGlyAlaProSerGlySerProSerProSerValSer 793  
DB 2266 GGGATCGTCTACGGCGACACCGCGGCTCTCGGACGCGCGC----- 2307  
QY 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSerProSer 813  
DB 2307 ----- 2307  
QY 814 ProSerProSerSerProSerProSerProSerProSerProSerProSerProSer 833  
DB 2307 ----- 2307  
QY 834 ProSerArgSerProSerProSerAlaSerProSerProSerProSerProSerProSer 853  
DB 2307 ----- 2307  
QY 854 SerSerProSerSerSerProSerProSerProSerProSerProSerGlyVal 873  
DB 2308 -----GGTACGAGCGCGCACCGACG-----GGCGCTGC 2337  
QY 874 LysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 893  
DB 2307 ----- 2307

```
Db 2338 ACGTGACATACAGGATCACGACCACTGGTGGCGGCTTCAG-----GCCGATGTA 2391
Qy 894 GlnValValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 913
Db 2392 CAGCTCCCAACACCGGCTGACCGCTGGGACGGCTGGTGGTGGG-----TGG 2442
Qy 914 PheThrArgAspGlySerSerThrLeuValThrAsnCysAspTrpAlaAlaIleGly 933
Db 2443 TCGTTCGGTGACGGGACGAGGAGTACCACAGCTGTGGAACGCTCTGTACGGCAGGCGGT 2502
Qy 934 CysGly-----AsnIleArgAlaSerPheGlySer 943
Db 2503 TCGGGGTGACCGCAGCAACCTGGCTGGAAACGGGAGGTGGCGCGGCTCC 2556

RESULT 5
US-10-026-994-4
; Sequence 4, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-4

Alignment Scores:
Pred. No.: 7.53e-105 Length: 2517
Score: 1631.50 Matches: 353
Percent Similarity: 57.2% Conservative: 130
Best Local Similarity: 41.8% Mismatches: 297
Query Match: 31.8% Indels: 64
DB: 5 Gaps: 23

US-09-917-376-1 (1-957) x US-10-026-994-4 (1-2517)
Qy 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44
Db 10 TCTCGAGTCTTGGCTTGTCTGGGGCGGCTC-----ATCCCTGCC 51
Qy 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGly 63
Db 52 CATGCTGCC-----TTTTCATGAAGAACGCTCAAGCTCGCGCGCGCGGCG 99
Qy 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83
Db 100 TTGCTCCCGCATCATCTTCATCCCAACAGAAAGCGGTAGCATATGACGACAGAT 159
Qy 84 IleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrp 103
Db 160 ATTGGGGGGCTGTACCGCTCAAC---GCCGACGACTCATGGACCGCGCTACCGATGGG 216
Qy 104 Val-----GlyTyrAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAla 119
Db 217 ATTGCTGATTAATGCCGCTGGCAACTGG-----GGCATCGACGCTGTGGCGCTT 267
Qy 120 AspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAsp 139
Db 268 GATCCGACGACCATCAAAAGGTATGTCGCGAGTCCGATGATATGAAACAGCTGGGAT 327
Qy 140 ProAsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrPro 159
Db 328 CCGAGTAATGGAGCCATCATCTCGTCTGTCAGACCGGCGCAACGCTGCTCTTCAACCA 387
```

```
Qy 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179
Db 388 TTGCCTTCAAAGTCGGGGTAAATGTCAGAGCGGAGCGGAGCGTCTGGCTGTC 447
Qy 180 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArg 199
Db 448 GATCCGGCACTCAACATCATCTACTTGTGCTCGCTCAGGAAACGGCTCTGGAAG 507
Qy 200 SerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyr 219
Db 508 TCTACGAGCGCGGCTGACCTTTTCAAGTCTGCTGTTCAACGGCACTGGAGCTAC 567
Qy 220 IleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal 239
Db 568 ATCCAGACCGGAGTGATTCCACCGCTTACAACAGCGACAGCAAGGAGTCTATGGGT 627
Qy 240 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 259
Db 628 ACGTTCGACTCAACAGCAGCAGCAGCGGGGAGCCAGTCTCGTATCTTTTGTGGCAG 687
Qy 260 AlaAspPro---AsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db 688 GCTGATAACATCACTGCTTCACTGATGTAGACGAATGCCGCTCCACGTGGAGTGT 747
Qy 279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db 748 GTACCGGGCAGCAGGAAATCTTCTCAACAGCGCAACTCAGCCAGCAGAGAG 807
Qy 299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGlyAspVal 318
Db 808 GCCTTGATCTGACCTATTCCGATGGCAGCGGCGGTATGATGGCACACTTGGCTCAGTG 867
Qy 319 TrpLysPheSerValThrSerGlyThrTrpArgIleSerProValProSerThrAsp 338
Db 868 TGGAGGTACGACATTGCGGGGAACTTGGAAAGACATCACCCCTGTCTCTGGATCAGAT 927
Qy 339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db 928 CTA-----TACTTTGGCTTTGGCGGCTTGGCTCGATTTGCAAGCAGCAGAAC 978
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db 979 CTTGTGTGTCTCTTTGAACTCTTGTGGCCAGATGCTCAGCTGTTTTCGGTCGACCGAC 1038
Qy 379 GlyGlyAlaThrTrpArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db 1039 TCTGGGCAACATGAGCCGATCTGGCGGTGGCGAGCTATCCGACTGAGACCTATTATAC 1098
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly---ValGlnProAsnPro 417
Db 1099 TACAGCATCTCACTCCCAAGCACCGTGGATCAAGAACAACTTTTATCGATGTGCGAGC 1158
Qy 418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433
Db 1159 GAGTCACCGTCCGATGCTCATCAAGCGCTCGGCTGGATGATGATGCTCTCCAGATT 1218
Qy 434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453
Db 1219 GACCAACCCGACAGCAACCACTGGCTCTACGCGCACCGGAATGACAACTTTTGGCGGCCAC 1278
Qy 454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeu 473
Db 1279 GATCTCAACATGGGACACGCGCCCACAATGTGTCAATCCCAATCACTGGCAGACGGCATC 1338
Qy 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAla 493
Db 1339 GAGGAATCTCCGTCAGGACCTGGCTCTGCACCCGCGGAAGGAGGACTATGTCGCCGA 1398
Qy 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510
Db 1399 GTCGAGACGACCAACGCTTTCACCTTTGGCCAGCAGAAACGACCTCGGGACATCGCGCAG 1458
```





Db 608 ATCCAGACCCGAGTGATTCCAAACGGCTACACAGCGACAGCAAGGACTCATGTGGGTT 667  
Qy 240 AlaPheAspLysSerSerSerLeuGlyGlnAlaSerThrIlePheValGlyVal 259  
Db 668 ACGTTGCACTCAACACGACGACGACGCGGGAGCAAGTCTCTATCTTTTGGCAGC 727  
Qy 260 AlaAspPro---AsnAsnProValPheThrPheArgAspGlyAlaThrTrpGlnAla 278  
Db 728 GCTGATAACATCACTGCTTCTATGTAGCAGCAATGCGGCTCCACGCTGAGTGCT 787  
Qy 279 ValProGlyAlaProThrGlyPheIleProHisGlyValPheAspProValAsnHis 298  
Db 788 GTACCGGGGACCCAGGAAATCTTCTCACAAGGCGAACTGACCGCAGCAGAGAG 847  
Qy 299 ValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyrllePheAspVal 318  
Db 848 GCCTTGATCTGACCTATTCGATGGCACAGGGCGGTATGATGCACACTTGGCTCAGTG 907  
Qy 319 TrpLysPheSerValThrSerGlyThrTrpThrArgLleSerProValProSerThrAsp 338  
Db 908 TGGAGGTACGACATTCGACAGGGGAACTTGGAAAGACATCAACCCCTGTCTCTGGATCAGAT 967  
Qy 339 ThrAlaAsnAspTyrllePheGlyTyrlleSerGlyLeuThrIleAspArgGlnHisProAsnThr 358  
Db 968 CTA-----TACTTTGGCTTTGGCGGCTTGGCTCGATTTGAAAGCCAGGAACC 1018  
Qy 359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIlePheArgSerThrAsp 378  
Db 1019 CTTGTTGTTGCTTTTCAACTCTTGGTGGCAGATGCTCAGCTGTTCTGGTTCACCGAC 1078  
Qy 379 GlyGlyAlaThrTrpThrArgLleTrpAspTrpThrSerTyrlleProAsnArgSerLeuArg 398  
Db 1079 TCTGGGACAAATGAGCGCCGATCTGGCGGTGGCGAGCTATCCGACTGAGACCTATTATC 1138  
Qy 399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly----ValGlnProAsnPro 417  
Db 1139 TACAGCATCTCAACTCCCAAGACCGGTGATCAAGAACAACTTTATCGATGTCAGCGAC 1198  
Qy 418 ProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAlaIle 433  
Db 1199 GAGTCACGCTCGATGTTCTCATCAGCGCTCGCTGGATGATGATGATGCTCTCGAGATT 1258  
Qy 434 AspProPheAsnSerAspArgMetLeuTyrlleGlyThrGlyAlaThrLeuTyrlleAlaThrAsn 453  
Db 1259 GACCAACCGACAGCAACCACTGCTCTACGCGACCGGAATGACAATCTTTTGGCGGCAC 1318  
Qy 454 AspLeuThrTyrlleAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeu 473  
Db 1319 GATCTCAACCACTGGACACGCGCCCAATGTGTCAATCCAACTCACTGGCAGACGGCATC 1378  
Qy 474 GluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAla 493  
Db 1379 GAGGAATCTCCGTCAGGACTGGCTCTCCACCGCGGAGGAGCTATTTGGCGCA 1438  
Qy 494 LeuGlyAspLeuGlyGlyPheThrHisAla-----AspValThrAlaValProSer 510  
Db 1439 GTCGAGACGACACGCGCTTCACTTTGCGACGAGAAACGACCTCGGGACATCCCGCAG 1498  
Qy 511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrlleAlaGluLeuAsn 530  
Db 1499 ACGGTCTGGGCAACGCCACATGGCCACCTCGACGCGGTGCTACTACCGCGGAATCG 1558  
Qy 531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550  
Db 1559 GTCAGAGCGTGTCCGCTCGGCAACACCGCGCGCACGCAACAG----- 1603  
Qy 551 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 570  
Db 1604 GTGGCCATCTCTCGACGCGCGCGGACGCTGGAGCATCGACTACGCGCGCGCACAGCTCC 1663  
Qy 571 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 590  
Db 1664 ATGAACGCGCGCACGCTGGCTATTTCGCGCGACGCGCACGATCTCTGTCTGTCAGCGCC 1723

Qy 591 AspProGlyGlnProValValTyrlleAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610  
Db 1724 TCGTCCGCGC-----GTGCGAGCGCTCGGAGTTTCCAGGCGAGCTTTGCTTCCTCTCG 1774  
Qy 611 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrlleAla 630  
Db 1775 AGCTCCCGCGCGCGCTCATCGCTCGAGAACCAACAGAGCTCTTCTTACGCGC 1834  
Qy 631 LeuSerAsnGlyThrPheTyrlleArgSerThrAspGlyGlyValThrPheGlnProValAla 650  
Db 1835 GGCTCCGATCGACTTTTACGTAGCAAGAGACACCGCAGCAGCTTC-----ACG 1885  
Qy 651 AlaGlyLeuProSerSerSerGlyAlaValGlyValMet-----PheHisAlaValProGly 668  
Db 1886 CGCGCG---CCCAAGCTGGCGCAGCGACGAGCATCGCGGATATCGCTGCTCACCGCACC 1942  
Qy 669 LysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrlleHisSerThrAsnGlyGly 688  
Db 1943 ACCGCGGCGACGTTGTATGTCTCGACCGACGCTCGGCATATTCGCTCCACAGCTCGGCGC 2002  
Qy 689 SerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707  
Db 2003 ACGACCTTTGGCCAGTCTCCACCGCTGACCAACCTACCATCGCTCGCTGGGTG 2062  
Qy 708 SerAlaProGlySerSerTyrlleAlaValPheValValGlyThrIleGlyValThr 727  
Db 2063 GGCTCA---GACTCGAAGCTGG---AACCTGTATGCTTCGGCACC-----GGCCCGTCA 2110  
Qy 728 GlyAla-----TyrlleArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAsp 745  
Db 2111 GGGGCTCGCTCTACCGCAGTGGAGACAGCGCGCTCTCGACGCGACATCCAGGGCTCC 2170  
Qy 746 GlnHisGlnTyrlleGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 765  
Db 2171 CAGGGCTTCGCTCATCGACAGCAAGTTCGCGCGCAGCGCGCAGCACCGCGCGGCAA 2230  
Qy 766 ValTyrlleGlyThrAsnGlyValGlyIleValTyrlle-----GlyAspIleGlyGlyAla 783  
Db 2231 GTCTACGTGGCACCAACCGCGCGGGGTCTTTTACGCTCAGGAGAACCGTTCGGCGCGCGC 2290  
Qy 784 ProSerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSer 803  
Db 2291 ACGGGCGGACT-----TCTCTGTCGACCAACGACAGCAGCAGC 2323  
Qy 804 ProSerProSerSerProSerProSerProSerProSerProSerProSerProSerSer 823  
Db 2324 AGCAGCAGTACTCTTCCGCGCAGCTCGAGCACACGCTGAGGTGAGGCTGTATCCACG 2383  
Qy 824 ProSerProSerProSerProSerProSerProSerProSerProSerProSerProSer 843  
Db 2384 ACCCGGGCTTCGAGCGTGACTTCTGTCGAGGACAGCTCGCGCGCGCTCCACCGGGTCA 2443  
Qy 843 rProSerPro 846  
Db 2444 GGGGTCCGCG 2453

## RESULT 7

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29



; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 1  
 ; LENGTH: 9025608  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (4187715)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-156-761-1

## Alignment Scores:

Pred. No.: 8,06e-97 Length: 9025608  
 Score: 1570.00 Matches: 393  
 Percent Similarity: 49.2% Conservative: 131  
 Best Local Similarity: 36.9% Mismatches: 346  
 Query Match: 30.6% Indels: 194  
 DB: 6 Gaps: 38

US-09-917-376-1 (1-957) x US-10-156-761-1 (1-9025608)

QY 4 SerGluAsnIleArgLeuThrMetArg-----SerArgArgLeuValSerLeu 19  
 DB 3162026 TCTCCAGAACGAGTATGTATGCGACACGCGCCCGGAGCAGACGAACCGTC----- 3161973  
 QY 20 LeuAlaIleThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIle 39  
 DB 3161972 CTCGCGGGACCGCGCG-----GCCGCGCGCTTCACGCGCGTCCCG-----GCC 3161928  
 QY 40 ThrAlaSerProAlaHisAlaAlaThrThr-----GlnProTyrThrThrSerAsn 56  
 DB 3161927 GTGGGTCCTCCAGGCGACCGCGCGAACCACCGCGCGCGCTGTACCGTGGCGCAAC 3161868  
 QY 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76  
 DB 3161867 GCCGTATCGGGGACCGGCTTCGTACCGCGGTCTCTCCACCCCTCGTACGCGGT 3161808  
 QY 77 IleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTTPAspAlaAlaAsnGlyArg 96  
 DB 3161807 TCGCCTACGCGCGGACCGACATCGGCGCGCTACCGTGGGACGACCGCGCGCGC 3161748  
 QY 97 TrpIleProLeuLeuAspTTPValGlyTrpAsnAsnTTPGlyTyrAsnGlyValValSer 116  
 DB 3161747 TGGACCCGCTCATCGACCACTCGGTGGGACGACTGGAACTCTCTCGCGTTCGAGCG 3161688  
 QY 117 IleAlaAlaAspProIleAsnThrAsnLysValTTPAlaAlaValGlyMetTyrThrAsn 136  
 DB 3161687 ATGGCGCTGACCCCGACCGACCGCGCGCGCTCTACCTCGCGTGGGACCTACGCCAG 3161628  
 QY 137 SerTTPAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTTPGln 156  
 DB 3161627 TCGTGGGCGGCAAC---GGCGCGTCTCGCTCGAGGACCGCGCGCGCACCTGGACC 3161571  
 QY 157 IleThrProLeuProPheLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArg 176  
 DB 3161570 CGCACCGACTGACCGTGAAGCTCGGCGGCAACGAGGACGCGCGCGCGCGTTCGAGCGA 3161511  
 QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyIysGly 196  
 DB 3161510 CTCCTGTGACCGCGCGACGACCGACCTCTGTGGTGGGACG---CGGACGACGCGG 3161454  
 QY 197 LeuTTPArgSerThrAspSerGlyAlaThrTTPSerGlnMetThrAsnPheProAspVal 216  
 DB 3161453 CTGCTCAAGTCGACCGACCGGGGCGCCACTTGGGACCGCGCGCGCTTCGCG----- 3161400  
 QY 217 GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlyVal 236  
 DB 3161399 -----CGGAAGGCGAACTCTCTCGCGGCGAGGAGTC 3161370

QY 237 ValTTPValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePhe 256  
 DB 3161369 GTGTTC-----CTCGTCGCCCGCGCGACCGTCTAC 3161337  
 QY 257 ValGlyValAlaAspProAsnAsnPro-----ValPheTTPSerArgAsp 271  
 DB 3161336 GCCGCTGGGTGACGGCGACCGCCTCGGGCAGCGGACCTGTACCGCAGCGCGAC 3161277  
 QY 272 GlyGlyAlaThrTTPGlnAlaValProGlyAlaProThrGlyPhe-----IlePro 288  
 DB 3161276 ---GGCACGACCTGGGGCGCGTCCCGCGCGCCCTCCGGCACCTCCCGCAAGTCCCG 3161220  
 QY 289 HisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGly 308  
 DB 3161219 CTCGCGCGCGCTACGACACGACACCGCGAGCTGTAGCTACGCGACGCGACCC 3161160  
 QY 309 GlyProTyrAspGlySerSerGlyAspValTTPLysPheSerValThrSerGlyThrTTP 328  
 DB 3161159 GGCCCGGGCGGCGAGTCCGACGCGAGCGTGCACAGCTCGTACCGCCACGCGGACGTGG 3161100  
 QY 329 ThrArgIleSerProVal-----ProSerThrAspThrAlaAsnAspTyrPhe 344  
 DB 3161099 ACCGAGTCCACCCCGTGAAGCGGGCGGACGACGAGCGCGCTCGCGCGACCTTC 3161040  
 QY 345 GlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGln 364  
 DB 3161039 GCCTACGGCGGGTCCGCTGACGCGCGCGCGCGCGCGCTCGTCTCGTCCACCAAC 3160980  
 QY 365 IleSerTTPProAspThrIleIlePheAspSerThrAspGlyGlyAlaThrTTPThr 384  
 DB 3160979 AACCGCTGGCGCGACGCGACCGTCTTCGCTCCACGCGCGCGCGCTACTCTGACG 3160920  
 QY 385 ArgIleTTPAspTTPThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSer 404  
 DB 3160919 TCCCTCAAGGAC-----GCCCGCGTGTTCGAGGTCC 3160887  
 QY 405 AlaGluProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeu 424  
 DB 3160886 GAGACTCTCTCTCGACTGGGCGCGACGAC-----AAGCGGAAGTTC 3160845  
 QY 425 GlyTTPMetAspGluAlaMetAlaIleAspPropheAsnSerAspArgMetLeuTyrGly 444  
 DB 3160844 GGCTGTGTGATCCAGCGCGCTCGCGTTCGACCGCTACGACTCCCGACGACGCTGTACGG 3160785  
 QY 445 ThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIle 464  
 DB 3160784 ACCGCGCGGACCTCTACGCGACCGCGACCTCAAGCGCTGG----- 3160743  
 QY 465 HisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPro 484  
 DB 3160742 -----GCACCGCGATCGCGGCTTGAGGAGAGCGCGCTGCGCCAACTGATCTGCC 3160689  
 QY 485 ProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAla 503  
 DB 3160688 CGGTGGGCGGACACCTGTATCAGCGACTCGGGGACATCGGTGTGTGTGTACCGAG 3160629  
 QY 504 AspValThrAlaValProSerThr---IlePheThrSerProValPheThrThrGlyThr 522  
 DB 3160628 CGGTTCAGCGGCTCTCCGTGCGCGCGCATCGGACGAAACCGCTGTTCGGGTGGCGAC 3160569  
 QY 523 SerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspPro 542  
 DB 3160568 GGACTCGCGGCGCGCGCGCGCGCTGTATGCTCGCGACGCGGTGGGCGGACCC 3160509  
 QY 543 SerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTP--- 561  
 DB 3160508 GGCAACGCG-----GGGTACTCCACGACGCGCGCGCGGACCTGGCG 3160467  
 QY 562 ---PheGlnGlySerGluProGlyValThrThrGlyGlyThrValAlaAlaSerAla 580  
 DB 3160466 CCCTTCGAGGCGCGCGCGCGACATCGCAAGGACGACCGGACCGCATCGCCACGAGTCC 3160407  
 QY 581 AspGlySerArgPheValTTPAla-----ProGlyAspProGlyGlnProVal 596

Db 3160406 GACGGGGCACACTGCTGTGGTCTCTGTCGACTGGGACGGC-----ACG 3160362  
Qy 597 ValTyrAla-----ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyVal 612  
Db 3160361 ACCTAGCGCGCCACCGCTCGACGGCAACGGCGGAGCTGGTCCGAGGTCTCTCTCTTC 3160302  
Qy 613 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProIysThrPheTyrAla----- 630  
Db 3160301 CCGAAGGGCGCCACACCGGTGCGCGACCCCGCGCATCCGACGGGTCTCTACGGCTACGAC 3160242  
Qy 631 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAla 650  
Db 3160241 TTGCAAAATGGAACTATACCGCAGCACTGACAGTGGCGGTCTGTTACGGCGCGGTGG 3160182  
Qy 651 AlaGlyLeuProSerSerGlyAlaValAlaGlyValMetPheHisAlaValProGlyLysGlu 670  
Db 3160181 GCGGAGTCTGCC-----TCCGGGACAGCAGTCTCAAGTGTGTCGGCGCGGACGAGC 3160125  
Qy 671 GlyAspLeuTrpIleAlaAlaSer---SerGlyLeuTyrHisSerThrAsnGlyGlySer 689  
Db 3160124 GCGCACTGTGGCTCTCGCGCAAATGGAACGGGTCTTACCGGTCCACCGACGGGGGAC 3160065  
Qy 690 SerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAla 709  
Db 3160064 ACCTTCGCCAGGATCGACAGCTGCTGGCTCTGACACCTCGGCTTCGCAAGGGGGC 3160005  
Qy 710 ProGlySerSerTrpProAlaValPheValGlyThrIleGlyGlyValThrGlyAla 729  
Db 3160004 GACGGCGCGACATACCGCGGATCTACAGGTGGTCTGACGGAGACCATCACCGCGCTC 3159945  
Qy 730 TyrArgSerAspAspCysGlyThrTrpValIleAsnAspAspGlnHisGlnTyr 749  
Db 3159944 TACCGTCCGACGACGCCCGCAGACATGGTTCGGATCAACGACGCGCCACCATGG 3159885  
Qy 750 GlyAsnTrpGlyClnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGly 769  
Db 3159884 GGGTGTATCGCGGAGCGCTCGTGGTGTACCGCGATCCACGCGCGGTCTACCTCGCC 3159825  
Qy 770 ThrAsnGlyArgGlyIleValTyrGlyAspIle-----GlyGlyAlaPro 784  
Db 3159824 ACCAAGCGCGCGCATCAGTACGGGGAGCGGTCTGATCCCGCGCTCACGGATGCCA 3159765  
Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804  
Db 3159764 CCGGGCGCGCATCTGTTCGTGGG---ACTACACCCCGAGCATGGCCCGCAGAGG 3159708  
Qy 805 Ser-----ProSerSerProSerPro----- 812  
Db 3159707 TCTGGCAGGAGGACGTACGCCTCATGAAGGAGCGCGGTCACTCCGTACCGTCCGGC 3159648  
Qy 813 ---SerProSerProSerPro----- 820  
Db 3159647 TCTTCTCTGGGCGGAGCTCGAACCCACCCCGGGGCGAGGAGTTCGGGTGCTCGACC 3159588  
Qy 821 -----SerSerSerProSerPro----- 828  
Db 3159587 GGCTATGACACGTATGACAGAACGGCATCGCGTCTGCTGGCCACCCCGCCGCT 3159528  
Qy 829 SerProSerPro----- 832  
Db 3159527 CGCCCCCGCTCGATGGGCGCTGCACCCCGAGACCTGCCCGCGAGGAGCGGCC 3159468  
Qy 833 -----SerProSerArgSerProSerProSerAla 842  
Db 3159467 GCATCGAGTGTGGGCGCGCGCAGACATTCGCCCACTCCAGTGGCGTCTACCGCGCT 3159408  
Qy 843 SerProSerProSerSerProSerProSerSer-----ProSerSer 858  
Db 3159407 ACGCGCGCGCCATCATCCGAGGACCTCGCGCGCTACCGCGGCCCATCCCGCTCACGA 3159348  
Qy 859 SerProSerProProSerProSerSerProValSerGlyGlyValLys----- 874

Db 3159347 TGTGGCACATCAACACAGTACTGACCTACGACTGGGGCGACAGGGCGGCCACCGCCT 3159288  
Qy 875 -----ValGlnTyrLysAsnAsnAspSerAlaProGlyAsp 886  
Db 3159287 TCCGCCGTGGTCTCAGGGCAGGTACGCACGCTCGACGACTCAACACGGCTGGG--- 3159231  
Qy 887 AsnGlnIleLysPro---GlyLeuGlnValValAsnThrGlySerSerValAspLeu 905  
Db 3159230 -----GCACGGCTTCTGGAGCCAGGCTACGACGACTGGGAGGACATCTCCCGCCG 3159177  
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925  
Db 3159176 GCCACGCCCATACATGAACACCCACCCAGGTCTCGACTCAAGGCTTCACTCCG 3159117  
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945  
Db 3159116 ACGCACTCATGGAGTGT---TCGTGCCGACGAGTACATCTCG--- 3159075  
Qy 946 ProAlaThrPro 949  
Db 3159074 CCGGCACACCC 3159063

RESULT 8  
US-10-156-761-2561  
; Sequence 2561, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2561  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2217)  
US-10-156-761-2561

Alignment Scores:  
Pred. No.: 1.33e-91 Length: 2217  
Score: 1442.00 Matches: 328  
Percent Similarity: 54.2% Conservative: 107  
Best Local Similarity: 40.8% Mismatches: 272  
Query Match: 28.1% Indels: 96  
DB: 6 Gaps: 25

US-09-917-376-1 (1-957) x US-10-156-761-2561 (1-2217)

Qy 7 IleArgLeuThrMetArgSerArgLeuValSerLeuLeuAlaAlaThrAlaSerPhe 26  
Db 1 ATGGCAGCGCCCGCGCGAGCAGACGACCGTCC-----CTCGCGGAGACCGCGC--- 51  
Qy 27 AlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAlaHisAla 46  
Db 52 -----GCCCGCGCGCTCACGGCTCCCC-----GCCGTGGGTCCCGAGCGCACGCC 99  
Qy 47 AlaThrThr-----GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGly 63  
Db 100 GCCGAACACCGCGCGCGCTCGTACCGTGGCGAACACCGCGTCTATCGGGGGACCGGC 159

QY 64 PheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAsp 83  
DB 160 TTCGTACCGGGTCTCTTCCACCCCTCCGTACCGGTCTCGCTACGCCGCGACCGAC 219  
QY 84 IleGlyGlyMetTyrArgTyrAspAlaAlaAsnGlyArgTyrIleProLeuLeuAspTyr 103  
DB 220 ATCGCGGGCGCTACCGCTGGGACACCGCGGCGCCGCTGGACCCGCTCATGACACAC 279  
QY 104 ValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 123  
DB 280 CTCGGCTGGGACGACTGAACTCTCTCGGGTTCGAGGCGATGGCTCGACCCACGACAC 339  
QY 124 ThrAsnLysValTyrAlaAlaValGlyMetTyrThrAsnSerTyrAspProAsnAspGly 143  
DB 340 CCGGACCGGCTTACTCTCGCGTGGGACACTACGCCAGTCTGTGGCGGGCAAC---GGC 396  
QY 144 AlaIleLeuArgSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPheLys 163  
DB 397 GCGGTCTCGCTCGAGGACCGCGCGCCACCTGGACCCGACCGACTGACCGTGAAG 456  
QY 164 LeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn 183  
DB 457 CTCGGCGGCAACGAGGACCGCGCGCGCGGTGAGCGACTCTCTGTCGACCCGCGCGAC 516  
QY 184 AspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSer 203  
DB 517 AGCGACACCTCTGCTGGGACG---CGGACGACGCGGCTGCTCAAGTCGACCGACCGG 573  
QY 204 GlyAlaThrTyrSerGlnMetThrAsnPheProAspValGlyTyrIleAlaAsnPro 223  
DB 574 GCGCGCACTTCGGGACCGCGGACCGCTTCCG----- 606  
QY 224 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLys 243  
DB 607 -----GCGAGCGGAATCTCTCCGGGAGGAGTCGTGTTCT----- 642  
QY 244 SerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn 263  
DB 643 -----CTCGTCGCGCGCGGCGCACCGCTACGCGGCTGGGGTGCACGCGCGAC 690  
QY 264 AsnPro-----ValPheTyrSerArgAspGlyGlyAlaThrTyrGlnAla 278  
DB 691 GGCACCTCGGCGACGCGGAACCTGTACCGCACGCGCGAC---GGCACGACCTGGGGGCG 747  
QY 279 ValProGlyAlaProThrGlyPhe-----IleProHisLysGlyValPheAspPro 295  
DB 748 GTCCCGCGCGCGCTCTCGGACCTCCGCCAAGTTCCTCGCGCGCGCGTACGACACG 807  
QY 296 ValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySer 315  
DB 808 CACACCGCGGAGCTGTAGTACGTACGGGACGACCCCGCGCGCGCGCGGCGGCGAGTCCGAC 867  
QY 316 GlyAspValTyrLysPheSerValThrSerGlyThrTyrThrArgIleSerProVal--- 334  
DB 868 GGCAGCGTGCAAGCTCGGTACCGCACGCGGAGCTGGACCGAGGTCAACCCGCGTGAAG 927  
QY 335 -----ProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 351  
DB 928 CCGGCGGGGACGAGGACGCGCTCGCGGACACCTTTCGCTACGCGGGGTCCCGCTC 987  
QY 352 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTyrTyrProAspThr 371  
DB 988 GACGCGCGCGCGCGCGCGCTCTGCTGTCCACCAACACCGCTGGCGCGCGCGCGAC 1047  
QY 372 IleIlePheArgSerThrAspGlyAlaThrTyrThrArgIleTyrAspTyrThrSer 391  
DB 1048 ACGGTCTTCGGTCCACGAGCGCGCGCGGTACCTGGAGCTCCCTCAAGAC----- 1098  
QY 392 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrLeuThrPhe 411  
DB 1099 -----GCGCGCGTGTTCGACGTTCGAGACTCCCTCTCTCGACTGG 1140

QY 412 GlyValGlnProAsnProValProSerProLysLeuGlyTyrMetAspGluAlaMet 431  
DB 1141 GCGGACGAC-----AAGCCGAAGTTCGGCTGGTGGATCCAGCGCGCTC 1182  
QY 432 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 451  
DB 1183 GCGGTGACCGCTACGACTCCGACGACGTCGTGTACGGGACCGCGGCGGACCTCTACGGC 1242  
QY 452 ThrAsnAspLeuThrLysTyrAspSerGlyGlyGlnIleHisIleAlaProMetValLys 471  
DB 1243 ACCCGGACCTCAAGCGCTGG-----GCACCGCGGATCCGC 1278  
QY 472 GlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGly---AlaProLeu 490  
DB 1279 GGCCTGGAGGAGAGCGCGCTGCCCAACTGATCTCGCCCGCGCGGGAGGACACCTG 1338  
QY 491 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 510  
DB 1339 ATCAGCGGACTCGGGACATCGGTGTGTGTACACGCGGCTCAGCGGTCTTCCGTCG 1398  
QY 511 Thr---IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeu 529  
DB 1399 CGCGGATCGCGACGAACCCGCTGTCGGTTCGGCGCGGACTCGCGACGCGCGCGCC 1458  
QY 530 AsnProSerIleLeuValArgAlaGlySerPheAspProSerSerGlnProAsnAspArg 549  
DB 1459 AGCGCGCGCTATGTCGTCCGACGCGCTGGGCGGACCAACGCAACGCGC----- 1506  
QY 550 HisValAlaPheSerThrAspGlyGlyLysAsnTyr-----PheGlnGlySerGluPro 567  
DB 1507 -----GCGTACTCCACGCGCGCGGACCTTGGCGCGCTTCGAGCGCCACCGCGAC 1560  
QY 568 GlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyr 587  
DB 1561 ATCGCAAGACGCGACCGGACCGATCGCCACCAAGTCCGACGCGCGGCGACACTGCTGTGG 1620  
QY 588 Ala-----ProGlyAspProGlyGlnProValValTyrAla----- 599  
DB 1621 TCCTTCTCGTCACTGGGACGCG-----ACGACGTACGCGGCGCCACCGCTCG 1665  
QY 600 ValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 619  
DB 1666 ACGGACAAACGCGCGGAGTGTCTCCGAGGTCTCTCTCTCCGGAAGGCGCGCACCGCGTC 1725  
QY 620 SerAspArgValAsnProLysThrPheTyrAla-----LeuSerAsnGlyThrPheTyr 637  
DB 1726 GCGACCGCGCGCGATCCGACGCGCTTCTACGCTTACGACTTCGACATGGAACGCTATAC 1785  
QY 638 ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly 657  
DB 1786 GCGACGACTGACAGTGGCGCTTCGTTACGCGCGCGCGCGGCGGACTGCCC---TCGCGC 1842  
QY 658 AlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTyrLeuAlaAla 677  
DB 1843 GACAGCCAGTTCAGAGTGTGTCGCGCGCGCGGACGAAGCGGACCTGTGGCTCTCCGCG 1902  
QY 678 Ser---SerGlyLeuTyrHisSerThrAsnGlyGlySerSerTyrSerAlaIleThrGly 696  
DB 1903 AAATGGAACGGGCTCTACCGGTCCACCGACGCGGCGGACCTTCGCGAGGACGACAGC 1962  
QY 697 ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAla 716  
DB 1963 TCCTGGGCTCTGACACCTCGGCTTCGGCAAGGCGCGCGACGCGCGCGGCGGCGGCG 2022  
QY 717 ValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736  
DB 2023 ATCTACGAGTTCGGCTCGAGGAGACCATCACCGCGCTCTACCGCTCCGACGACGCGCGC 2082  
QY 737 ThrThrTyrValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTyrGlyGlnAlaIle 756  
DB 2083 AGGACATGGGTCCGGATCAACGACGCGCCACCACTGGGGGTGGATCGCGAGGCGGTC 2142  
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776

```

Db      2143 GTGGGTGACCCGCGCATCCACGGCCGGTCTACTCCCAACACGGCGCGCATCCAG 2202
Qy      777 TyrGlyAsp 779
Db      2203 TACGGGGAG 2211

RESULT 9
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Ramseier, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2668)
US-09-927-827-21

Alignment Scores:
Pred. No.: 4,18e-72 Length: 3668
Score: 1167.50 Matches: 299
Percent Similarity: 48.6% Conservative: 137
Best Local Similarity: 33.3% Mismatches: 351
Query Match: 22.7% Indels: 112
DB: 3 Gaps: 28

US-09-917-376-1 (1-957) x US-09-927-827-21 (1-3668)
Qy      23 ThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer 42
Db      1136 ACCACAAACAGGACATGGTGGTGTCTTCTGAGCGCTGCTG---CTGTTGCTCTTTTCGACG 1192
Qy      43 Pro-----AlaHisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAla 58
Db      1193 CCAGCGTGGTCCGCGGAGCCGCCACGTCGCGGCCCTTACCAGTGGCGAGGTGCGCC 1252
Qy      59 IleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGlyIleLeu 78
Db      1253 ATTGGCGGTGGCGCTTTGTACCGGTGTGCTGTTTCATCCCGCGCAACGTGTGCTGGCC 1312
Qy      79 TyrValArgThrAspIleGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIle 98
Db      1313 TATCGCGCACCATGTGGTGGCGGTACCGCTGGGATCGCAGCGCGCAGCAGTGGACC 1372
Qy      99 ProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAla 118
Db      1373 GCGCTGACCGACTGGTGGCGCTGCTGACGACTGAACTGTATGGGCAATCGCATTCGCC 1432
Qy      119 AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 138
Db      1433 GTCCACCCGCGCATGCCGATGGCTGTATCTGGCGCGCGCACCTATATGCATGAA--- 1489
Qy      139 AspProAsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThr 158
Db      1490 CGCGCGCGCACTGCCGACAGTTGGCTGCTCTCAACCGCGCGCGCAGTTTCGAGCGTGC 1549
Qy      159 ProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAla 178
Db      1550 GACCTCCGCTTTAAGCTGGTGGTGAACACAGCTGGCGCGCCCAATGCGCAGCGCTGGCG 1609
Qy      179 ValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrp 198

```

```

Db      1610 GTGACCCGCGCATGGCGCGTGTCTGTCTGGCTCGCGCATGCC---GGCCTGTGG 1666
Qy      199 ArgSerThrAspSerGlyValaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 218
Db      1667 CGTAGCGACATCGCGCGCGCATGGCGGAGGTGGCTGTTTCGCGACCGCGCGTG 1726
Qy      219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 238
Db      1727 GCCGTGCCACCGCGCGCAATCATGTTGGCGCGAGCAGCGGTG---GGGATCGCCTTT 1783
Qy      239 ValAlaPheAspLysSerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGly 258
Db      1784 GTCGTGTTTCGACGCGCAGTGGCAACACGCTCGCAACACCGCGCATCTACGTGGGC 1843
Qy      259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAla 278
Db      1844 GTGTCCACCGCAACACAGCAGCTGTATGTGTCGGAAGATGCCGCGCGCATTTGGGCACCG 1903
Qy      279 ValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHis 298
Db      1904 GTGCCCGGCAACACCGCGGCTGCGCCGAGCCACATGGCGCGGCGCAGCATGGGCAC 1963
Qy      299 ValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspVal 318
Db      1964 ---TGGTATCTGAGCTATGGCGACCGCGCGCGGACCTGATGCGCGGGGAGCCTTG 2020
Qy      319 TrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAsp 338
Db      2021 TGGAAATTCACGCGCGCACAGGCGCGTGGGTGAGATCAGCCCGATTCGCGAG---CCA 2077
Qy      339 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThr 358
Db      2078 GCCAGTGGCGATGATTCGGCTGGGTGGGTGGCGGTGGATCCGCAACATCCGACAGTG 2137
Qy      359 IleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAsp 378
Db      2138 CTGCTGGCGCAGCACCTTCGCGCGTCGCGACGCGCGCGCAGCTGTATCGACGCGTGAT 2197
Qy      379 GlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArg 398
Db      2198 GGTGGCAAGCATCGACCGCGCTG---TTGGCCGAT 2230
Qy      399 TyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProPro 418
Db      2231 CGCGTGTTCGATCAGCGCGCGCGCGTGG---ACCGCACATGCCACGCGCGCAC--- 2281
Qy      419 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 438
Db      2282 -----TGGATGGG---GGCTGGCGCATCGATCCGTTCGACGCG 2317
Qy      439 AspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 458
Db      2318 AACCATGCGCTGTTCGTACCGCGTACCGCATCTGGGCTCGGCGCTCGCGCAATCTG--- 2368
Qy      459 AspSerGlyGlyGlnIleHisIleAlaPro-----MetVal 470
Db      2369 -----CAGGATTCGCGCGCACCGCGCGCGCTGCTGCTGCTGCTGCTGCCAGAC 2416
Qy      471 LysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu 490
Db      2417 CGTGGCTGGGAAGAACCGTCCGCTGAGCTGTCTACCGCATGCTGGTGGCGCGCATCTG 2476
Qy      491 IleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSer 510
Db      2477 CTCAGCGCTCGCGCATATCGACGGCTTCGCGCATGACGACCTGACCGCGCTG---CAG 2533
Qy      511 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 530
Db      2534 TTGCAGTACGCGCGCGCGCGCTGACCAATGGCAAGCATCGATCGCGCGCGCGCGCG 2593
Qy      531 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 550

```

```
Db 2594 CCGCAGTGGGTGGTGGCGAGCGGTACCGTGGCGCGCCCGCCGCAACCAACGAAATCCGCGCG 2653
Qy 550 sValalaPheSerThrAspGlyGlyLysAsnTrpPheGln---GlySerGluProGlyGlu 569
Db 2654 C---TGTATTACGGCTGATGCGCGGCAAGCAATGCGTTCGTTGCGAGTGAGCGCCGACG 2710
Qy 569 yValThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPr 589
Db 2711 CGGGCAGGCGCCGCGACATTCGATTCGCGCATGCGCGCAGGAGGTGGTGGCGCAC 2770
Qy 589 oGlyAspProGlyGlnProValValTrpAlaValGlyPheGlyAsnSerTrpAlaAlase 609
Db 2771 GCGCGCGCGCGCAAT-----TGGCGCAGCTCCGACTTCGCGCGCAGTGGCAGCGTGT 2824
Qy 609 rGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValaAsnProTysThrPheTy 629
Db 2825 GAACGCCCTGCGGAACCCGCGTGGTGGTACCGATCGGTGGAGCAGCAGCGCGCTGGTA 2884
Qy 629 rAla-----LeuSerAsnGlyThrPheTyArgSerThrAspGlyGlyValThrPheGlu 647
Db 2885 TCGCGTGGATGTCGCCAGCGCGGAGCTGTACGAGAGCACCAGTCCGCGCGCAGTTTCCG 2944
Qy 647 nProValAlaAlaGlyLeu----- 653
Db 2945 T-----GCGACCGGTGTACAGGTGGCGAGCCCGCGCGCGATGAGCGCACCCGCGCGCA 2998
Qy 654 -----ProSerSerGlyAlaValAlaGlyValMetPheHisAlaValProGlyLysGluGl 671
Db 2999 GCTGCGTCCGAGCCGCGGCGCGAGGGTGTATCTGCGCCAGCCCGCGCAG----- 3053
Qy 671 yAspLeuTriLeuAlaAlaSerSerGlyLeuTyHisSerThrAsnGlyGlySerTr 691
Db 3054 -----GGTGTGATCGCTGGCGAGCAGGT-----GCGCT 3082
Qy 691 pSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGl 711
Db 3083 GCAGTGCTCTCACAGCGCGGAGCAGCGTCTCGTGGCATCGCAAGCGCGTGGCGG 3142
Qy 711 ySerSerTyProAlaValPheValGlyThrIleGlyValThrGlyAlaTyAr 731
Db 3143 CGTGGCGCGCGCGCTGTATCTGCGCGCGCGTGCAGGCGGTGATGGCGTTCG 3202
Qy 731 gSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyGlyAs 751
Db 3203 CTCGAGCAGTGGCGCGCGCAGTGGCAGCGCATCACGATCAGCGCACCGCTTCGCGG 3262
Qy 751 nTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyIleGlyThrAs 771
Db 3263 C---CCGTACAGCGTGCAGCGGTGATCCGCGCATTCGCGCGGTGTACTTCGCGCACCG 3319
Qy 771 nGlyArgGlyIleValTyGlyAspIleGlyAlaProSerGlySerProSerProse 791
Db 3320 CGGCGCGCGCATTTTTCACGCGCATCCGAGATGACGATGCATCCCTGTCGCGA----- 3374
Qy 791 rValSerProSerAlaSerProSerLeuSerProSerProSerProSerSerPro-- 810
Db 3375 ---CGTCTGCTGCTTACCGTGTACGCGCGCGCGCGCTGGCGCGCTGGCGCGCGCGTG 3430
Qy 811 -----SerProSerProSerProSerSerProSerSerProSerProSer-- 827
Db 3431 ACACGCGGACCTGGAGATCACCGCATGACCCCGT-CCCACATTTCCCGCGCTGCTTG 3489
Qy 828 -----ProSerProSerProSerProSerProSerProSerProSerProSerPro 836
Db 3490 CTGGGTGTGTGTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3549
Qy 837 -----SerProSerProSerAlaSerProSerProSerProSerProSerProSer 850
Db 3550 GCCGATGGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3609
Qy 850 oSerProSerSer-----SerProSerSerSerProSerProSerProSerPro 864
Db 3610 GCGCGCGCAATCACGTTGAGTTTCGATGCGCAAGCAGCGCGCGCGCGCGCGCGCGCG 3661
```

## RESULT 10

```
US-10-395-241-11
; Sequence 11, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; FILE REFERENCE: 073756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 2646
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(2558)
US-10-395-241-11
```

```
Alignment Scores:
Pred. No.: 2,646-69 Length: 2646
Score: 1125.50 Matches: 287
Percent Similarity: 49.8% Conservative: 131
Best Local Similarity: 34.2% Mismatches: 294
Query Match: 21.9% Indels: 127
DB: 7 Gaps: 32
```

US-09-917-376-1 (1-957) x US-10-395-241-11 (1-2646)

```
Qy 17 ValSerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaAlaLeuGlyValLeuPro 36
Db 114 TTAGATAGTGGTGGCAGTCACCTCCCTCGGGAAGCGCTTACTGCCCTTTTCGATTCTG--- 170
Qy 37 IleAlaIleThrAlaSerProAlaHisAlaAlaThrGlnProTyThrTrpSerAsn 56
Db 171 -----GCGTCGCTCGCGTGGCGGAGGAGCAGCTACGAGTTTCAAGAAT 212
Qy 57 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 76
Db 213 GTCGCGATCGCGCGCGCGGTACATTACCGGGATGTGCGGACCCCAAGACCAAGGAC 272
Qy 77 IleLeuTyValArgTrpAspIleGlyMetTyArgTrpAspAlaAlaAsnGlyArg 96
Db 273 CTGCTGTCGCGCGCGCAGCAGCATTTGGCGCGGTACCGTGGGACGCGCAGCTCCAG 332
Qy 97 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyAsnGlyValValSer 116
Db 333 TGGATCCCGCTCAACGACTTTATCGAGGCGCAGGACATCAACATTATGGCACCAGTCTG 392
Qy 117 IleAlaAlaAspProIleAsnThrAsnLysValTrpAlaValGlyMetTyThr--- 135
Db 393 ATCGCGCTGGAGCCCAACACCCCGAGCGGTGTACTCTCGCAGCGCGCGCTATGTGCGC 452
Qy 136 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 155
Db 453 GACGAGTGG-----GCGCGTTCATGTGTCTCCAGACCGCGCGCGCGTCTGT 500
Qy 156 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu 175
Db 501 ACAATCTACGAGTCCCGCTTCCGATGGGCGCAGCAGCATGGGAGCGCAACAATGGCGAG 560
Qy 176 ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyThrPheGlyAlaProSerGlyLys 195
Db 561 CGCCTCGCTGTCAACCCCGTTCAACTCGAACGAGGTCTGGATGGGTACGGGTACA---GAG 617
Qy 196 GlyLeuTyArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215
```





|  |    |     |   |     |
|--|----|-----|---|-----|
|  | Qy | 52  | TyrThrTrpSerAsnValAlaIleGlyGlyGluArgPheValAaspGlyIleValPheAsn | 71  |
|  | Dd | 13  | TACGAGTTTCAAGAATGTCCGATCGCGCGCGGTACATTACCGGATTTGTTCGGCAC      | 72  |
|  | Qy | 72  | GluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyMetTyArgTTPasp       | 91  |
|  | Dd | 73  | CCAAAGACCAGGACTCTCTGACGCGGCACGACATTCGCGCGGTACCGCTCGGAC        | 132 |
|  | Qy | 92  | AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTy   | 111 |
|  | Dd | 133 | GCAGCCAGTCCAAGTGATCCCGCTCAACACTTTATCGAGCGCAGGACATGAACATT      | 192 |
|  | Qy | 112 | AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaIaVal   | 131 |
|  | Dd | 193 | ATGGGCCAGCATGTCATCGCTGACCCCCAACAAACCCGACAGCGTGTACCTCGCGCAG    | 252 |
|  | Qy | 132 | GlyMetTyThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp   | 150 |
|  | Dd | 253 | GGCGCTATGTCGGCAGCAGTGG-----GCGCGCTTCTATGTCCGAAGAC             | 300 |
|  | Qy | 151 | GlnGlyAlaThrTrpGlnIleThrProLeuProPhelLysLeuGlyGlyAsnMetProGly | 170 |
|  | Dd | 301 | CGCGCCAGTCGTTTACAATCTACAGATCGCGCTCCGATGGCGCCACGACATGGGA       | 360 |
|  | Qy | 171 | ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspIleLeuTyPheGly      | 190 |
|  | Dd | 361 | CGCAACAATGGCGAGCGCTCGCTGTCAACCCGTTCAACTCGAACGAGGTTCGATGGGT    | 420 |
|  | Qy | 191 | AlaProSerGlyLysGlyLeuTyArgSerThrAspSerGlyAlaThrTrpSerGlnMet   | 210 |
|  | Dd | 421 | ACCGGTACA--GAGGGTATCTGAAGAGTTCGACCGCCGACACCTCGACAACAGCTC      | 477 |
|  | Qy | 211 | ThrAsnPheProAspValGlyThrTyIleAlaAsnProThrAspThrThrGlyTrpGln   | 230 |
|  | Dd | 478 | AGTCCNTCCGAGCGCTTC-----ACCAACGGTATCGGATACACG                  | 519 |
|  | Qy | 231 | SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGln     | 250 |
|  | Dd | 520 | TCG-----GTCAATTCGACCCC-----GAA                                | 540 |
|  | Qy | 251 | AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg  | 270 |
|  | Dd | 541 | CGTAATGCGACATCTACGCGAGCGGACTGCCCCGAGGCG---ATGTACGTCACGCAC     | 597 |
|  | Qy | 271 | AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle-----      | 287 |
|  | Dd | 598 | GACGCGGTGTCTCGTGGAGCCAGTGGCGGCGCAGCGCTCCAGCTGGCTCAACAGGACC    | 657 |
|  | Qy | 288 | -----ProHisLysGlyVal  | 292 |
|  | Dd | 658 | ACGGCGGGTTCGCGACAAGAAGCCGCGCTCGATCGCGCGGACGCCCATGAAGTCGCT     | 717 |

```

Qy 611 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 624
Db 1750 ---GTTCCCGCTGGCGACCTGAAGCCAGAGCTGCTCAATGCTCTCGGCAAGTCCAG 1806
Qy 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 644
Db 1807 GATGGCAGTTCTACGTTACGATGCGCGCAAGTTCTTCGTCACGACGCGGGGAG 1866
Qy 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664
Db 1867 TCGTATCCGCAAGAGCGCGCGACTTGTCACT-----GACACATCGCTCATGCTGCC 1920
Qy 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 683
Db 1921 GTGAACCCCTGGGTGGCGCGGAGCTGTGGTGTCTGTCCGAGGCGGTCTCTTCCAC 1980
Qy 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702
Db 1981 TCGACCGACTTGGGGCTGTTACGAGGTAGTACCGCAACGCGACCTCGTGAGC 2040
Qy 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 718
Db 2041 GTCGGCGCCCAAGTCCAAAGTCGACGCGCAAGAAGGTAGCGCGCTCGCGGTCTTC 2100
Qy 719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736
Db 2101 ATTCGGGGCACCGCAAGCTGGAAGCGACATCGGCTGTACCGCTCGGACGACACGCG 2160
Qy 737 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756
Db 2161 AGCACTGGACCGCGCTCAATGACGAGGACCAACTACTCGGGC---CCACCATGATC 2217
Qy 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyVal 776
Db 2218 GAGCGCCAGCCCAAGGCTACCGCGCGTGTATAGGACGACGCGCGGTATCGTG 2277
Qy 777 TyrGlyAspIleGlyAlaProSerGlySerProSerProSerProSerProSerAla 796
Db 2278 TACCGCGACCTTACCAACAAGAGAGC---AACGAGGAGAGTTCGACGCGCAAGTGGCGC 2334
Qy 797 Ser-----ProSerLeuSer 801
Db 2335 AACGGCCAGAGGCGACGACTGTATGTGAAAAGAGAGTCTGGGTACCTGGTGCCAC 2394
Qy 802 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 821
Db 2395 GCGGTTCCATGGCTGATATCGATCCGNAATTCGAGTCCGTCGACAGCTTGGCGCGCA 2454
Qy 821 rSerSer-ProSerProSerPro 828
Db 2455 CTCGAGCACCACCACCACCA 2477

RESULT 12
US-10-395-241-13
; Sequence 13, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME
; FILE REFERENCE: Q73756
; CURRENT APPLICATION NUMBER: US/10/395,241
; CURRENT FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: JP 2002-83433
; PRIOR FILING DATE: 2002-03-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 2367
; TYPE: DNA
; ORGANISM: Geotrichum sp. M128
; FEATURE:

```

```

; NAME/KEY: mat_peptide
; LOCATION: (1)..(2367)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2367)
US-10-395-241-13

Alignment Scores:
Pred. No.: 9,28e-69 Length: 2367
Score: 1117.00 Matches: 281
Percent Similarity: 50.4% Conservative: 124
Best Local Similarity: 35.0% Mismatches: 279
Query Match: 21.8% Indels: 120
DB: 7 Gaps: 31

US-09-917-376-1 (1-957) x US-10-395-241-13 (1-2367)
Qy 52 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 71
Db 10 TACGAGTTCAAGAATGTCCGATCGCGCGCGGGGTACATTACCGGGATTGTCCGCGAC 69
Qy 72 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 91
Db 70 CCAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGGCGCGGTACCGCTGGAC 129
Qy 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111
Db 130 GCAGCACGTCCAAGTGGATCCCGCTCAACGACTTTATCGAGCGCGCAGCACATGAACATT 189
Qy 112 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131
Db 190 ATGGCACCGATCGATCGCTGGACCCCAACCAACCCGACAGGCTGTACCTCGCGCAG 249
Qy 132 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 150
Db 250 GGGCGCTATGTGCGGACGAGTGG-----GCGCGCTTCTATGTGTCGGAAGAC 297
Qy 151 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuGlyGlyAsnMetProGly 170
Db 298 CGCGCCAGTGGTTTACATCTACGAGTCGCGCTCCCGATGGCGCCCAACGACATGGGA 357
Qy 171 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 190
Db 358 CGCAACAATGGCGAGCGCTCGCTGTCAACCGTTCAACTCGAACGAGGCTCGATGGGT 417
Qy 191 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 210
Db 418 ACGGTATACA---GAGGGTATCTGGAAGAGTTCGGACCGCGCCCAAGACCTGGACAACGTC 474
Qy 211 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 230
Db 475 ACGTCCATCCGCGACGCGTTC-----ACCAACGGTATCGGATACAG 516
Qy 231 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 250
Db 517 TCG-----GTCAATTTTCGACCCC-----GAA 537
Qy 251 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 270
Db 538 CGTAATGTCACCATCTACGAGCGCGCTGTCGCGCGAGGCG---ATGTACGTCACGCGAC 594
Qy 271 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 287
Db 595 GACGCGGTGTCTCTGTTGGAGCCAGTGGCGGCGCCAGCTCCAGCTGGCTCAACAGGACC 654
Qy 288 -----ProHisLysGlyVal 292
Db 655 ACGGGCGGCTCCCGGCAAGAGCCGCGTGTATCGCGCGCGCCGACCCATGAAGTCGCT 714
Qy 293 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 312
Db 715 CTCACCCCC-----AACTTCTCTACGTGACTTACGCGCGACTACCTCGTCCATGGGGC 768

```



QY 313 GlySerSerGlyAspValTyrPheSerValThrSerGlyThrTyrThrArgile--- 331  
Db 769 GTACAGTTCCGGAAGTCTGGCCGAGAACCGACCTCGGGCGCTCGGAGACATTA 828  
QY 332 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 345  
Db 829 CCCCGCGTGGCAACTCGTCGCTGCGGTACAACACAGAGCGTTCCCTCGGGCGGA 888  
QY 346 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 365  
Db 889 TTTTCGGTCTCAGCGTCGACGGCGAGCAACCAACCGCTCGTCTCATCACC---CTC 945  
QY 366 SerTyrTyrProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 382  
Db 946 GACCGGACCGCGGACCGCCCTCGACAGCATCTACCTCTCAACCGATGCCGGCGGAC 1005  
QY 383 TrpThrArgIleTyrAspTyrThrSer-----TyrPro 393  
Db 1006 TGAAGAGACGTACCCAGCTCTGTCCCGTCCAACTCGAAGTAACCTGGGGCCACCCG 1065  
QY 394 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrPheGlyVal 413  
Db 1066 ACTAAGCGCGCGGTAC---AAGGACGGCACGCTGTTCGTGGCTCGACTTCAACAAC 1122  
QY 414 GlnPro-----AsnProValProSerPro-----LysLeuGly 425  
Db 1123 GGTCCCAAGTGGGGGGATACGGTGGCGGACGAGTACGCGCGGCTCACCAAGTTTGGC 1182  
QY 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
Db 1183 TGTGGATGAGCGCTGTCTTATCGATCCGTTCACCCGACGACCTGTATGACGGCAG 1242  
QY 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 461  
Db 1243 GGGGGACCATCTGGGCGACCGACACGCTCTCCGTGCGAAGGACTGG----- 1293  
QY 462 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 477  
Db 1294 -----GGCGCGAGCTGGTACCTCCAGATCGAGCGGTATCGAGAGAATCG 1338  
QY 478 ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 497  
Db 1339 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGCTCTCTGTGGGATCGGTGACATT 1398  
QY 498 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 517  
Db 1399 AGCGCATGAAGCAGCAGCAGCTCACCAAG-----CCCCAAGATGTTTGGTGGCCCGC 1455  
QY 518 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 537  
Db 1456 TTCTCCAACTCGACAGATCGAGCTCGGGCAACTTCCCAAGCTGTGTCTCGCGCC 1515  
QY 538 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 557  
Db 1516 GGATCCTCGGACACGAGTACGACGCGGTGGCGCGGTGGCTACGCGACTGACGCG 1575  
QY 558 GlyLysAsnTyr-----PheGlnGlySerGluProGlyGlyValThr----- 572  
Db 1576 GGAGCGGTGGACATCTTCCCTACCTCCCTCGCTGCAATGAACGGCGCACTTACCAG 1635  
QY 573 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTyrAlaPro-----Gly 590  
Db 1636 GCGACGATGATGATCGACGCGGAGCGGAGCAGATCGTGTGTGTCACCAAGCTTGCAC 1695  
QY 591 AspProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 610  
Db 1696 GAGCAGGCTCGGACCGTGTGTACTCGCAACGACTATGCAAGACGTGTCT----- 1746  
QY 611 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 624  
Db 1747 ---GTTCCCGTGGCGACCTGAAGGCCCGCAGACTGCCAATGTGCTCTCGCAAGGTCCAG 1803  
QY 625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644

Db 1804 GATGGCACGTTCTACCTTACCGATGGCGCAAGTTCTTCTCTCGACCGAGCGGGAAG 1863  
QY 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 664  
Db 1864 TCGTATGCGCCCAAGGGCGCGGACTTGTCACT-----GGCACATCGCTCATGCTGCC 1917  
QY 665 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 683  
Db 1918 GTGAACCCCTGGGTGGCGCGGACGCTCTGGTGGCTGTTCGCCGAGGCGGTCTCTTCCAC 1977  
QY 684 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 702  
Db 1978 TCGACCGACTTGGGCGCTCTGTCACGAGGTAGGTACCGCAACCGACCCCTCGTGAGC 2037  
QY 703 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 718  
Db 2038 GTCCGGGCCCCAAGTCCAACTCGGACGCAAGAGCTAGCGGCCCTCCGGGTCTTC 2097  
QY 719 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 736  
Db 2098 ATCTGGGCGCACGCAAGGCTTGAAGCGACATCGGCTGTACCGCTCCGACGACACCGC 2157  
QY 737 ThrThrTyrValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 756  
Db 2158 AGCACCTGACGCGGTCAATGACGAGGACACACTACTCGGGC---CCCACCATGATC 2214  
QY 757 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 776  
Db 2215 GAGCGCGACCCCAAGTCTTACGGCGGTGTATCTAGGCAACGCGCGGTATCGTG 2274  
QY 777 TyrGlyAspIle 780  
Db 2275 TACGCGGACCTT 2286  
RESULT 13  
US-10-420-191-1  
; Sequence 1, Application US/10420191  
; Publication No. US20040067569A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; FILE REFERENCE: 10210.200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; CURRENT FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; PRIOR FILING DATE: 2002-04-19  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 1  
; LENGTH: 5698  
; TYPE: DNA  
; ORGANISM: Tricoderma reesei  
US-10-420-191-1  
Alignment Scores:  
Pred. No.: 6, 19e-56 Length: 5698  
Score: 940.00 Matches: 374  
Percent Similarity: 30.1% Conservatives: 136  
Best Local Similarity: 22.1% Mismatches: 315  
Query Match: 18.3% Indels: 871  
DB: 7 Gaps: 59  
US-09-917-376-1 (1-957) x US-10-420-191-1 (1-5698)  
QY 25 SerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSerProAla 44  
Db 128 TCTCGAGTCTTCCCTTGTCTCTGGGGCGGTC-----ATCCCTGCG 169

45 HisAlaAlaThrThrGln----- 50  
170 CATGCTGCCCTTCTCCAGAGAGCTCAGGAACGGAAACAGGACCCCGGAGTAGGGACGG 229  
51 -----ProTyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGly 67  
230 GTACGCGGGAATTCATGGAGAGAGCTCAAGCTCGGCGGGCGGGCGGCTTCTCCCGCGC 289  
68 IleValPheAsn-----GluGlyAlaPro----- 75  
290 ATCATCTCTCCAAAGTACCTTCTTGCACTTCGAGCGCGCGCGCGCGAAGCAGGGCGCG 349  
76 -----GlyIleLeuTyrValArgThrAspIleGlyGlyWet 87  
350 TAGTAGAAGGTTCCCAAGACAAAGCGTAGCATATGCAGCAACAGATATTTGGCGGCGTG 409  
88 TyrArgTrp----- 90  
410 TACGGCTCAAGGGTCTGTTTCCGCATCGTATACGTGCTTGCTATTAACCGCCGAC 469  
91 -----AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal----- 104  
470 ATGGCGGAGTTCGCGAGAGCTCATGACCGCGCTCACGGATGGGATTTGCTGATAATGCC 529  
105 GlyTrp-----AsnAsnTrpGly----- 110  
530 GGCTGGCACAAGCGGCTGCTGAGTACCTGGCGGCGAGTCCCTAACGACTATTACGG 589  
111 -----TyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleVal 127  
590 CGACCGTGTCTGGGCGATCGAGCTGTGCGCTTGTATCCGAGGACGATCAAAAGGTG 649  
128 TrpAlaAlaVal----- 131  
650 TATGCGCGAGTAGCCCGTAGTCGACAAAGCGAGTACGCTGCTAGTTTCCAC 709  
132 -----GlyMetTyrThrAsnSerTrpAsp----- 139  
710 ATACGGCGTTCACGGCATGTATACGAACAGCTGGTC-TGTGATGCTCTCAGATCTAGACC 768  
139 ----- 139  
769 TATGATTGGACGGCGGTACATATGCTTTGTCGACCAACACTACAGAGAGTCTAGATCTGG 828  
140 -----ProAsnAspGlyAlaIleLeuArg 147  
829 ATACTAACCTGCTGACATTGGCCATATATAGGATCCGAGTAAATGAGCATCATTCGC 888  
148 SerSerAspGln----- 151  
889 TCCTCAGACCGGACTGTAAACCGGTATATATCCCTAGGCTCATTTACCTCGGTAGTAAGCG 948  
152 -----GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn 167  
949 AGCAGTCTGGCGGGCGCAACGTGCTCTTCCACCAACTTGGCTTCAAGTGGGGGTAAAC 1008  
168 MetProGlyArg----- 171  
1009 ATCCAGAGCGCCCGCTTCCACAGGAAGTGGTTGAACGGGAAGTTTACGCCCCCATNTG 1068  
172 -----GlyMetGlyGlyArgLeuAlaValAspProAsnAsnAspAsnIleLeu 187  
1069 TACGGTCTCGGGAGCGCGAGAGCGTCTGGCTGTGATCCGGCCAACTCCACATCATC 1128  
188 TyrPheGlyAlaPro----- 192  
1129 TACTTTGGTCTCTCGGCCCTCTCGCAGACCGACAGCTAGGCGCGTTGAGTTGTAGTAG 1188  
193 -----SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 207  
1189 ATGAACACCAACGACGCTCAGGAACCGCCCTCTGGAAGTCTACGACGCGCGGTGACCTTT 1248  
208 SerGlnMet----- 210

1249 TCCAAGGTCTCGGCGAGTCTTTTCCCGGAGACCTTTCAGATGCTTCCGCGCAGCTGGAAA 1308  
211 -----ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThr 227  
1309 AGGTTCCAGAGCTGTTTCCCGCAACTGGGAGGTACATCCAGACCCGAGTGTATCCAAC 1368  
228 GlyTyrGln----- 230  
1369 GGCTACACAGCAGCAAGTGCCTTGCCTGCATGTAGGGTCTGGGCTCACTAAGGTTG 1428  
231 -----SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 247  
1429 CCGATGTTGTCCGACAAAGAGACTCATGTGGTTAGCTTTCAGTCAACAGCAGCAGC 1488  
248 LeuGlyGlnAla----- 251  
1489 ACCGGGGAGCCCTGTTGTTCTTGAGTACACCAACGAGCTGAGTTGGTGGTCTGC 1548  
252 -----SerLysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPhe 267  
1549 TGGCCCCCTCGGACGCTCTCGTATCTTTTGTGGCAGGCTGATAACATCACTGCTTCAGTC 1608  
268 Trp----- 268  
1609 TATGTGACACGTGCAGAGCATAGAAACACCGTCCGACTATTGTAGTGAAGAGTCAAG 1668  
269 -----SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 286  
1669 ATACACTCGTCAATCGCGCTCCACGTGGAGTGTCTGTACCGGGCAGCCAGGAAATAC 1728  
287 IleProHis----- 289  
1729 TTTCTCACAAGTTACGGCGAGGTGCACCTCAGACATGCCCCGTCGTCCTTATG 1788  
290 LysGlyValPhe-----AspProValAsnHisValLeuTyrIleAlaThrSerAsn 306  
1789 AAAGAGTGTTCGGGAAACTGCAGCCAGCAGAGAAGGCTTGTATCTACCTATTCCGAT 1848  
307 ThrGlyGlyPro----- 310  
1849 GGCACAGGCGCGCGCTTTGACGTGCTGCTCTCTCCGAAACATAGACTGGATAAGCTTA 1908  
311 -----TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 326  
1909 CCGTGTCCCGCTATGATGGCACACTTGGCTCAGTGTGGAGGTACACATTCAGGGGGA 1968  
327 ThrTrpThrArgIleSerProValProSerThrAspThr----- 339  
1969 ACTTGGAAAGACATACTACCGTGTGAACGAGTCAACCTCCATGCTGTAACGTCCCCCT 2028  
340 -----AlaAsnAspTyrPheGlyTyrSerGlyLeu 349  
2029 TGAACCTTTCTGATCACCCTGCTCTCTGGATCAGATCTATATCTTTGGCTTTGGCGCCT 2088  
350 ThrIleAsp----- 352  
2089 GGCCTCGATTGTAGTGGGACAGAGACCTTAGTATAGAAACCGAAACCGCGGAA 2148  
353 -----ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpPro 369  
2149 CCGAGCTAAACCAAGCCAGGAACCTTGTGTGTCTTCTTCAACTCTTGGTGCCA 2208  
370 AspThr----- 371  
2209 GATGCTCAGTGTGTTTTCGTCCTTGGGAACAACAACGAAGAACTTGAGAACCACCGGT 2268  
372 -----IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp 389  
2269 CTACAGTTCGACTTTCGTCGACCGACTCTGGGCAACACATGAGCCCATCTGGCGGTGG 2328  
390 ThrSerTyrProAsnArgSer---LeuArgTyrValLeuAspIleSerAlaGluPro--- 407

|      |    |   |      |
|------|----|---|------|
| 2329 | Db | CGGAGCTATCCGAAAGCCAGCTGGCTGAGACC--CTGTTGTACTCTCGGGCTAGACCCGCGAC | 238  |
| 407  | Qy | -----   | 407  |
| 2388 | Db | CCGCTCGATAGGCACCTGAGACCTATTACTACAGCATCTCACTGAGTCACCTCTTAAACGATC | 2447 |
| 408  | Qy | -----TrpLeuThrPhe-----GlyValGlnProAsn-----                      | 416  |
| 2448 | Db | CGATCGAATGGCTGACTCTCGGATAATGATGTCTGAGAGTCACCTCACTGAGGAATTCGTAG  | 2507 |
| 417  | Qy | -----ProProValPro-----  | 420  |
| 2508 | Db | GCTACGCTTACCGACTGACTCTGCTTCAAGACTCCCAAGACCCGTGGATCAAGAAACAAC    | 2567 |
| 420  | Qy | -----   | 420  |
| 2568 | Db | TTTATCGATGATGACTGAGACGAAAGTTCTGAGGGTTTCGTGGCACCCTAGTTCCTTGTGTTG | 2627 |
| 421  | Qy | -----SerPro-----LysLeuGlyTrpMet                                 | 427  |
| 2628 | Db | AAATAGCTACATCGAGCGAGTCACCGTCCGATGGTCTCATCAAGCGCCTCGCGCTGGATG    | 2687 |
| 428  | Qy | AspGluAlaMetala-----  | 432  |
| 2688 | Db | ATTGAGTCTCTCGGCTCGCTCAGTGGCAGGCTACAGAGTAGTTTCGGGAGCGCGACTTAC    | 2747 |
| 433  | Qy | -----IleAspProPheAsnSerAspArgMetLeuTyTrGlyThrGlyAla             | 447  |
| 2748 | Db | TAACTCAGACAGAGATTGACCCACCGACAGCAACCATGGGCTCTACGGCACCCGGAATG     | 2807 |
| 448  | Qy | ThrLeuTyTrAlaThr-----   | 452  |
| 2808 | Db | ACAATCTTTGGCGTCTAACTGGGTTGGCTGTGTTGGTGACCGAGATGCGGTGGCCTTAC     | 2867 |
| 453  | Qy | -----AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAla              | 467  |
| 2868 | Db | TGTTAGAAACCGGGCCACGATCTCACCAACTGGGACACGCGCCACAATGTGTCAATCCAA    | 2927 |
| 468  | Qy | ProMetValLysGly-----  | 472  |
| 2928 | Db | TCACTGGCAGACGGGTGCTAGAGTGTTGACCCGTGCGCGGTGTACACATTAGGTT         | 2987 |
| 473  | Qy | -----LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGly              | 487  |
| 2988 | Db | AGTGACCGTCTGCGCATCGAGGAATCTCCGTCCAGGACTTGGCCTCTGCACCCGCGGA      | 3047 |
| 488  | Qy | AlaProLeuIleSer-----  | 492  |
| 3048 | Db | AGCGAGTATTGGCGTAGTCTCTTAAGAGCAGGTCTCTGGACCGGAGACGTGGGCGCGCT     | 3107 |
| 493  | Qy | -----AlaLeuGlyAspLeuGlyGlyPheThrHisIleAlaAspValThrAla           | 507  |
| 3108 | Db | TCGCTCGATTAACCCCGCAGTCGGAGACGACAACGGGCTTCACCTTTCGACGAGAAACGAC   | 3167 |
| 508  | Qy | Val-----  | 508  |
| 3168 | Db | CTCGGACATCGCGCGCTCAGCCTCTGCTTTCGCCGAATGGGAAAACGGTCGCTTTGCTG     | 3227 |
| 509  | Qy | -----ProSerThrIlePheThrSerProValPheThrThrGlyThrSerVal           | 524  |
| 3228 | Db | GAGCCCTGTAGCGGACGAGCGGTCTGGGCAACGCCCAACATGGGCCACCTCGACGAGCGTC   | 3287 |
| 525  | Qy | AspTyTrAla-----   | 527  |
| 3288 | Db | GACTACCGCGGGAGCGTCTGCCAGACCCGTTGCGGGGTATCCGGTGGAGCTGCTCGCAG     | 3347 |
| 528  | Qy | -----GluLeuAsnProSerIleIleValArgAlaGlySer-----                  | 539  |
| 3348 | Db | CTGATCGGCGCCTACTCTGGTCAAGAGCGGTCTCCGGTCTGGGCAACCCCGCGCAGCGAA    | 3407 |
| 540  | Qy | -----PheAspProSerSerGlnProAsnAspArgHisValAla                    | 552  |
| 3408 | Db | CAGTGCGCCATCTTGAGCCAGATTCTCGCAGACGGCGCAGCGTGTGGCGCGCGTTCGTT     | 3467 |

|    |      |  |      |
|----|------|--|------|
| Qy | 553  | Phe-----SerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGly             | 569  |
| Db | 3468 | GTCACCGGTAGACTCTCCGACGGCGGCACGTCGAGCATCGACTACGCGGCCACACG             | 3527 |
| Qy | 570  | ValThrThrGlyGly-----   | 574  |
| Db | 3528 | TCCATGAACCGCGCAGCGTCCCGCGCGCTGCACCTCGTAGTGATGCCCGCGGTGTC             | 3587 |
| Qy | 575  | -----ThrValAlaAlaSerAlaAspGlySerArgPheValTPrpAlaPro                  | 589  |
| Db | 3588 | AGGTACTTGGCGCACGGTGGCTATTCCGCCGACGCGCACGATCTCTCTGGTCGACC             | 3647 |
| Qy | 590  | GlyAspProGly-----  | 593  |
| Db | 3648 | GCCTCGTCGGCGGTGCCACCGGATAAGCGGCTGCCGTCTGTAGGAGACCACTGG               | 3707 |
| Qy | 594  | -----GlnProValValTyraValAlaGlyPheGlyAsnSerTrpAlaAlaSerGlnGly         | 611  |
| Db | 3708 | CGGAGCAGGCGCTGCACGCGCTCG--CAGTTCCAGGGCAGCTTGTCTCGTCTCGAGC            | 3764 |
| Qy | 612  | ValProAlaAsnAlaGln-----  | 617  |
| Db | 3765 | CTGCCCGCGGCGCGA <sup>CT</sup> CGCGAGCGTCAAGGTCCTCGAAACGGAGGAGAGACTCG | 3824 |
| Qy | 618  | -----IleArgSerAspArgValAsnProIylsThrPheTyraAlaLeu                    | 631  |
| Db | 3825 | GACGGCGCGCGGCTCATCGCTCGGACAAGAAGAACACAGCGTCTCTACGCGCGC               | 3884 |
| Qy | 632  | SerAsnGlyThrPhe-----   | 636  |
| Db | 3885 | TCGGATCGACCTTTAGTAGCGGAGCCTCTTCTTGGTTGTCGAGAAGATCGCGCGC              | 3944 |
| Qy | 637  | -----TyrArgSerThrAspGlyValThrPheGlnProValAlaAla                      | 651  |
| Db | 3945 | AGGCCTAGCTGGAAAAACGTCCAGCAAGGACACCGGACAGACTTC-----ACGCGC             | 3995 |
| Qy | 652  | GlyLeuProSerSerGlyAlaValGlyValMet-----                               | 662  |
| Db | 3996 | GGG---CCCAGCTGGGACGCGAGGATCGACTGTTCTGTGGCGGTCTGGAAGTGC               | 4052 |
| Qy | 663  | -----PheHisAlaValProGlyLys   | 669  |
| Db | 4053 | CGCCCCGGGTTCACCGTCGCGTCCCTCGATCCGGATATCGTCTACCCGACCAACC              | 4112 |
| Qy | 670  | GluGlyAspLeuTrpLeuAlaAlaSerSer-----                                  | 679  |
| Db | 4113 | CGCGGCACGTTGTATGTCTCGACCGAGCTAGGCCCTATAGCAGAGTGGCTGGTGG              | 4172 |
| Qy | 680  | -----GlyLeuTyHisSerThrAsnGlyGlySer                                   | 689  |
| Db | 4173 | CGCCCGTGCACATACAGACTGGTGTCTCGCATATTCGCTCCAAGACTCGGCGACG              | 4232 |
| Qy | 690  | Ser-----   | 690  |
| Db | 4233 | ACCTTTGGCCAAGTCTCCACCGCCTGAAGCCGTATAAGSGAGGTGTCTGAGCCCGTGC           | 4292 |
| Qy | 691  | -----TrpSerAlaIleThrGlyValSerSerAlaValIleValGly                      | 704  |
| Db | 4293 | TGGAACCCGGTTCAGAGGTGGCGGAGCTCCAACACCTTACCAGATCCCGCTGGTGGG            | 4352 |
| Qy | 705  | PheGlyLysSer-----AlaPro  | 710  |
| Db | 4353 | TCAGGCTCGAACTGGAACTGATGCTGCTGTTGTGGATGTTAGCGGGAGCCACACCCG            | 4412 |
| Qy | 711  | GlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla---         | 729  |
| Db | 4413 | AGTCCGAGCTTGACCTTGGACATACGATCGGCACC-----GGCCCGTCAGGGGTCTCG           | 4466 |
| Qy | 730  | ---TyrArgSerAspAspCysGlyThrTrpVal-----                               | 740  |
| Db | 4467 | CTCTAGCCAGCTGGAGACAGCGCGCTCTCTGGAAGCCGTGGCGGGCAGTCCCCGAGCG           | 4526 |

Qy 741 -----LeuIleAsnAspGlnHisGlnTyrGly--- 750  
 Db 4527 GAGATGCGGTCACTCTGTGCGCGGAGGACCTCGGACATCCAGGCGCTCCAGGGGTTTC 4586  
 Qy 750 ----- 750  
 Db 4587 GGCTCCATCGACAGCACCAAGGTCCGCGGAGCGGCTGTAGTCCCGAGGGTCCCGAAG 4646  
 Qy 751 -----AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 766  
 Db 4647 CCGAGGTAGTGTCTGTTCAGCGCGCTCGCGCAGCACCGCGGG-----CAAGTC 4700  
 Qy 767 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp----- 779  
 Db 4701 TACGTGGGCACCAACGCGCGGGGCTCTTTTACGCTCAGGCGGTGTGGCGCGCGCTTCAG 4760  
 Qy 780 -----IleGlyGlyAlaPro 784  
 Db 4761 ATGACCCGTGGTTCGCGCGCGCGCAGAAATGCGAGTCCGACCGTCCGCGCGGCGC--- 4817  
 Qy 785 SerGlySerProSerProSerValSerProSerAlaSerProSerLeuSerProSerPro 804  
 Db 4818 ACGGGGGGACTTCTCGTCGACCAACGACGAGCAGCAGTACTTGGCAGCGCGCGC 4877  
 Qy 805 SerPro-----SerSerSerProSerProSerProSerPro 816  
 Db 4878 TGCCCGCCCTGAAGGAGCAGTGTCTCGTCTCGTCTCGTCTCTCTCGCAGCTCG 4937  
 Qy 817 SerSerSerProSerSerProSerProSerPro----- 828  
 Db 4938 AGCACCACGCTGAGGTGAGCGTGTATCCACGACCGCGGCTTGGAGAGCGCGGTGAGC 4997  
 Qy 829 -----SerProSerProSer----- 833  
 Db 4998 TCGTGGTGCAGCTCCAGTCCACATAGTGTGGGCGCGGACGAGCGTGTCTGTCG 5057  
 Qy 834 -----ProSerArgSerProSerProSerAlaSerProSerProSerSerProSerPr 852  
 Db 5058 AGGACAGCTCGGCGCGCGCTCCACCGGGTTCAGGGGTGCGCGCTGCCACTGAAGCAGC 5117  
 Qy 852 oSerSerProSerSerProSerProSerProSerProSerProSerProSerProSer 865  
 Db 5118 TCCTGTGTCAGCGCGCGGCGGCGGAGGTGCCCGCCAGTCCCGCAGC 5157

## RESULT 14

US-09-917-384-2  
 ; Sequence 2, Application US/09917384  
 ; Publication No. US20030096342A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: DECKER, STEPHEN R.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOSE TOLERANT  
 ; FILE REFERENCE: 40170.6US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,384  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 3687  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 ; OTHER INFORMATION: GuxA  
 US-09-917-384-2

## Alignment Scores:

Pred. No.: 6.43e-48 Length: 3687  
 Score: 823.00 Matches: 251

Percent Similarity: 47.1% Conservative: 46  
 Best Local Similarity: 39.8% Mismatches: 173  
 Query Match: 16.0% Indels: 164  
 DB: 3 Gaps: 24  
 US-09-917-376-1 (1-957) x US-09-917-384-2 (1-3687)  
 Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445  
 Db 307 TGGATGACCCGATCGCTGCGATCAACGGCGTCAACGGC-----GGAACC 351  
 Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHis 465  
 Db 352 GGCTTGACGACATATCTCGACCGCCCTCTCCAG-----CAGCAGGGA 396  
 Qy 466 IleAlaProMetValLysGlyLeuGluGlnThrAlaValAsnAspLeuSerProPro 485  
 Db 397 ACACCCCTGAAGTC-----ATTGAGATTGTCTACTACGATCTG-----CGGGA 441  
 Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 505  
 Db 442 CGCGACTGCGCGGCGCTCGCTCCAAACGCGAAGTCCCGCT---ACGGCAGCAGGTTTG 498  
 Qy 506 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 525  
 Db 499 CAGACCTAT---GAAACGCGAGTACATCGATCCGATTGCGAGTATCTCGACATCCGAAG 555  
 Qy 526 TyrAlaGluLeuAsnProSerIleIleValAlaGlySerPheAspProSerSerGln 545  
 Db 556 TACTCCAGCTGCGGATCGTGCAGCATCAIT-----GAGCCGACTCGCTG 600  
 Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565  
 Db 601 CCAACGC-GGTAC-----CAATAT 620  
 Qy 566 GlnProGlyGlyValThrThrGlyGlyThrValAla-----AlaSerAlaAspGly 582  
 Db 621 GAGCATTCACAGCGTGTCAACGCGGCGTCCGCTATT-ACGAGCAGGCGATCGAGTACGCGC 679  
 Qy 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592  
 Db 680 TCAGAAATTGACGCCCATTCGAAACGTTGATCATCTACATGAGCGCCGCCCTCCGCT 739  
 Qy 593 -----GlyGlnProValVal-----TyrAlaValGlyPheGlyAsnSerTrp 606  
 Db 740 GGCTTGGTGGCCCAATATCCAGCGGATACGTACAGAAAGTCCAGAGGTCTCTCAACG 799  
 Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624  
 Db 800 CGACATCGGGTCAACGGCATCGACGCTTCGTCACCAACACACGCGGATACACGCGCT 859  
 Qy 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637  
 Db 860 TGAAGGAGCGCTTCATGACCGCCACCCAGCAGGTGCGCGGACAGCGGTGAGTGGCGGA 919  
 Qy 638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648  
 Db 920 ATTCTACAGTGAATCTCTGACATC-GACGAAGCCGACTACGCGGTGACTGTGACTCG 978  
 Qy 649 -----ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666  
 Db 979 CGGCTCGCGCGCTGGCTTCCAGAGCAGC-----ATCGCGATGCTCATCGACCTTA 1032  
 Qy 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 686  
 Db 1033 -----CGCAAC 1038  
 Qy 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703  
 Db 1039 GGTGGGGTGTCTCCGAGCAGCAACACGAGCCGCGGACCGGATGTCAACACCTTC 1098  
 Qy 704 -----Gly 704

```

Db 1099 GTCAACAGTGAAGATTGACCTTCGGCAGCAGCCGCGCTGTGTGTGCAACAGACGCT 1158
Qy 705 PheGlyysrAlaProGlySerTyrProAlaValPhe----- 718
Db 1159 GCGGCGCTCGCCAGCGCGCAGCAGCGCGAGCGCGAGCTTCCGAGCGCCACCTCGAC 1218
Qy 719 -----ValValGlyThrIleGlyGlyValThrGlyAla----- 729
Db 1219 CCGTATGTCTGGATCAAGCCCGCGGTGAGTCGGACGCGCAAGCGTGGAGCGATCCG 1278
Qy 730 -----TyrArgSerAsp-----CysGlyThrThrTrpValLeuIleAsnAsp 745
Db 1279 ACAACTGGCAAGAGTCGACCCCATGTGCGACCGAGTACACGCG----- 1326
Qy 746 GlnHisGlnTyrGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGTACTGACCAAGCGGTACCGAACTCCCGATCGCGCGCAGTGG 1380
Qy 761 AlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 780
Db 1381 TTCGCGCGCAGTGTGACAGCTTGTGCGGAACGACGCGCGAGCGGTGCGACGTCGACC 1440
Qy 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCGCGCGCTCCGCCCGCGAGTCGCGTGGCTTCGCCAGTCCGAGCCGAGT 1500
Qy 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCGCGAGCAGCTCGCCATCGCGTCCGCTCGCGTCTCGAGCTCGAGCCGCTCGGTCG 1560
Qy 814 Pro-----SerProSerSerProSerProSerProSer 825
Db 1561 CCGAGCGCGAGTCCGAGCGCGAGTAGCTCGCGTCCGCTCGCGTCTCGAGCTCGAGCCGCT 1620
Qy 826 ProSerProSerProSerProSerProSerProSerProSerProSerProSer 845
Db 1621 CCGTCCGCGAGCGCGAGTCCGAGCGCGAGTAGCTCGCGTCCGCTCGCGTCTCGAGCTCGAGC 1680
Qy 846 ProSerSerProSerProSerSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCGTCCGCGAGCGCGAGTCCGAGCGCGAGTAGCTCGCGTCCGCTCGCGTCTCGAGC 1740
Qy 866 SerSerProValSerGlyValLysValLysValGlnTyrLysAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCCGCGGTGTGCGGTGGCGTGAAGTGCAGTACAGAACATGATTCGCGCGCGCGGT 1800
Qy 886 AsnAsnGlnLysProGlyLeuGlnValValAsnThrGlySerSerValAspLeu 905
Db 1801 GATAACAGATCAACCGCGGTCTCCAGTTGGTGAATACCGGGTCTGTCGCGTGGATTG 1860
Qy 906 SerThrValThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 1861 TCGACGGTACGGTGGCGGTACTGGTTTCCCGGGATGGTGGTGGTGGTGGTGGTGGTGGTGG 1920
Qy 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
Db 1921 AACTGTGCTGGCGGGGATGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1980
Qy 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956
Db 1981 CCGCGCGAGCGCGCGCGGACACCTACTCTGCGAG 2013

```

RESULT 15

```

US-09-917-383-2
; Sequence 2, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS

```

```

; TITLE OF INVENTION: CELLULOSE
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1.
; SEQ ID NO 2
; LENGTH: 3687
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
US-09-917-383-2

```

```

Alignment Scores:
Pred. No.: 6,43e-48 Length: 3687
Score: 823.00 Matches: 251
Percent Similarity: 47.1% Conservatives: 46
Best Local Similarity: 39.8% Mismatches: 173
Query Match: 16.0% Indels: 164
DB: 3 Gaps: 24

```

US-09-917-376-1 (1-957) x US-09-917-383-2 (1-3687)

```

Qy 426 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 445
Db 307 TGGATGGACCGCATCGCTCGATCAACGGCGCTCAACGGC-----GGACCC 351
Qy 446 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHis 465
Db 352 GCGTTTGACGACATATCTGGACCGCGCTCTCCCCAG-----CAGCAGGGA 396
Qy 466 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 485
Db 397 ACCACCCCTGAAGTC-----ATTGAGATTGTCATCTACGATCTG-----CCGGA 441
Qy 486 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspVal 505
Db 442 CCGGACTCGCGCGCGCTCGCTCCAAACGGCGAAGTCCCGCT---ACGGCAGCAGGTTG 498
Qy 506 ThrAlaValProSerThrPheThrSerProValPheThrThrGlyThrSerValAsp 525
Db 499 CAGACCTAT---GAAACCGCAGTACATCGATCGATTCGATTCGAGTATCTCGACAAATCCGAG 555
Qy 526 TyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGln 545
Db 556 TACTCCAGCCTCGCGATCGTGCAGATCATT-----GAGCCGAGCTCGCTG 600
Qy 546 ProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySer 565
Db 601 CCAACAGC-GGTCAC-----CAATAT 620
Qy 566 GluProGlyGlyValThrThrGlyGlyThrValAla-----AlaSerAlaAspGly 582
Db 621 GAGCATTCAGCGTGTGCAACCGCGTCCGCTATT-ACGAGCAGGATCGATCGACGCGC 679
Qy 583 SerArg-----PheValTrpAlaProGlyAspPro--- 592
Db 680 TCACGAAATTCACGCCCATTCGGAACGCTGTACATCTACATGACGCGCGCCACCTCGGCT 739
Qy 593 -----GlyGlnProVal-----TyrAlaValGlyPheGlyAsnSerTrp 606
Db 740 GCGTTGGTGGCCCAATAATATCCAGCGGATACGAGGAGTCCAGAGTCTCTCAACG 799
Qy 607 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn----- 624
Db 800 CGAGCATCGGGTCAACCGCATCGCGCTTCGTCCACCAACACCGCGGAATTACACGCCGT 859
Qy 625 -----ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr 637
Db 860 TGAAGGAGCGGTTCATGACCGCCACCGCAGGTCGCGCGGACGAGCCGCGTGGAGTCGCGCA 919

```

```
QY 638 ArgSerThrAspGlyGlyValThrPheGlnPro----- 648
Db 920 ATTCTACCAAGTGAATCCCTGACATC-GACGAAGCCGACTACGCGGTGACTGTACTCG 978
QY 649 -----ValAlaLaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 666
Db 979 CGGCTCGTCCCGCTGCTTCCAGAGC-----ATCGCATGCTCATCGACACCTTA 1032
QY 667 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTrpHisSerThrAsn 686
Db 1033 -----CGCAAC 1038
QY 687 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal----- 703
Db 1039 GGTGGGGTGGTCCGAACGAACCAACAGGCCGAGCAGCCGCGACCGCATGTCAACACCTTC 1098
QY 704 -----Gly 704
Db 1099 GTCAACAGTCGAAGATTGACCTTCGGCAGCAGCCGCGCTGTGGTGAACCAACGACGGT 1158
QY 705 PheGlyLysSerAlaProGlySerSerTy-ProAlaValPhe----- 718
Db 1159 GCGGGCTCGCCAGCCGCGCAGCGCAAGCCCGAGCGACTTCCGAACGCGCACCCTCGAC 1218
QY 719 -----ValValGlyThrIleGlyValThrGlyAla----- 729
Db 1219 GCGTATGCTCTGGATCAAGCCGCGGTGAGTCGGCAGCGCACAGCGCTGCGAGCGATCCG 1278
QY 730 -----TyrArgSerAspAsp---CysGlyThrThrTrpValLeuIleAsnAspAsp 745
Db 1279 ACACTGGCAAGAGTCGGACCCCATGTGCGACCCGACCGTACACGACG----- 1326
QY 746 GlnHisGlnTrpGlyAsnTrpGlyGlnAla-----IleThrGlyAspHis 760
Db 1327 -----TCGTACGGGGTACTGACCAACCGGTACCGAACTCCCGATCGCGGCCAGTGG 1380
QY 761 AlaAsnLeuArgValThrIleGlyThrAsnGlyArgGlyIleValThrGlyAspIle 780
Db 1381 TTCGCGCGCGAGTTTGACCACTGTTCGCGAACGACGCGCGCGGTGCCGAGCTCGAC 1440
QY 781 GlyGly-----AlaProSerGlySerProSerProSerValSer 793
Db 1441 AGCTCGAGCCCGCGCTCGCGCGCGAGTCCGTGCGCTTCGCGAGTCCGAGCCCGAGT 1500
QY 794 ProSerAlaSerProSerLeuSerProSerProSerProSerProSerProSer 813
Db 1501 CCGAGCCCGAGCAGCTCGCCATCGCCGTCGCGTCCGAGTCCGAGCCCGCTCCCGTCG 1560
QY 814 Pro-----SerProSerSerSerProSerSerProSer 825
Db 1561 CCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGCTCCGAGCTCGAGCCCGTCT 1620
QY 826 ProSerProSerProSerProSerProSerProSerProSerProSerAlaSerProSer 845
Db 1621 CCGTCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGCTCCGAGCTCGAGC 1680
QY 846 ProSerSerSerProSerProSerProSerProSerProSerProSerProSerProSer 865
Db 1681 CCGTCTCGGTCGCGAGCCCGAGTCCGAGCCCGAGTAGCTCGCGCTCGCGCTCCCGAGC 1740
QY 866 SerSerProValSerGlyGlyValIysValGlnTrpLysAsnAsnAspSerAlaProGly 885
Db 1741 TCGTCGCGGTGTTCGGGTGGGTGAGGTGAGTACGATCAAGAACAAATGATTGGCGCCGGT 1800
QY 886 AspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerSerValAspLeu 905
Db 1801 GATAACAGATCAACCGGGTCTCCAGTTGTGTGAATACCGGGTCTGTCGGTGTATTG 1860
QY 906 SerThrValThrValArgTrpTrpPheThrArgAspGlyGlySerSerThrLeuValTyr 925
Db 1861 TCGACGGTGACGGTGGCTGTGTTTCCACCGGGATGGTGGTCTCGACACTGGTGTAC 1920
QY 926 AsnCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 945
```

Search completed: March 2, 2006, 17:16:06  
Job time : 15122 secs

Db 1921 AACTGTGACTGGGCGCGCATGGGGTGTGGGAATATCCCGCCCTCGTTCGGCTCGGTGAAC 1980  
QY 946 ProAlaThrProThrAlaAspThrTyrLeuGln 956  
Db 1981 CCGGCGACCGCGACGGCGGACACCTACTCTGCAG 2013

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocollaboration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:24:59 ; Search time 1026.96 Seconds

(without alignments)  
2043.020 Million cell updates/sec

Title: US-09-917-376-1

Perfect score: 5134

Sequence: 1 MDRENTRLTWSRLVLL.....RASFGSVNPATPTADTYLQX 957

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 7218512 seqs, 1096188070 residues

Total number of hits satisfying chosen parameters: 14437024

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer p2n.model -DEV=xlp  
-O=/abss/ABSSWEB.spool/US09917376/runat\_02032006\_091502\_8293/app\_query.fasta.1  
-DB=Published Applications NA New -QFMT=fastep -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs03p  
-USER=US09917376 @CGN\_1\_1067@runat\_02032006\_091502\_8293 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7.  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New:

1: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Length | DB ID | Description                             |
|------------|-------|--------------|-------|---|
| 1          | 282.5 | 5.5          | 9903  | 12 US-11-052-554A-517 Sequence 517, App |
| 2          | 273   | 5.3          | 9474  | 12 US-11-052-554A-526 Sequence 526, App |
| 3          | 273   | 5.3          | 11151 | 12 US-11-052-554A-525 Sequence 525, App |
| 4          | 270.5 | 5.3          | 6615  | 12 US-11-052-554A-518 Sequence 518, App |

|   |       |       |       |   |
|---|-------|-------|-------|---|
| 5 | 264.5 | 5.2   | 6210  | 12 US-11-056-470-1 Sequence 1, Appli          |
| 6 | 264.5 | 5.2   | 7572  | 12 US-11-052-554A-527 Sequence 527, App       |
| 7 | 261.5 | 5.1   | 70513 | 8 US-10-995-561-13368 Sequence 13368, A       |
| c | 8     | 260   | 5.1   | 4146 12 US-11-052-554A-522 Sequence 522, App  |
| c | 9     | 258.5 | 5.0   | 88421 12 US-11-052-554A-520 Sequence 520, App |
| c | 10    | 250   | 4.9   | 3950 12 US-11-052-554A-520 Sequence 520, App  |
| c | 11    | 250   | 4.9   | 114801 12 US-11-121-086-22 Sequence 22, Appl  |
| c | 12    | 247.5 | 4.8   | 3240 12 US-11-052-554A-529 Sequence 229, App  |
| c | 13    | 247.5 | 4.8   | 7231 12 US-11-136-527-2622 Sequence 2622, Ap  |
| c | 14    | 247   | 4.8   | 6360 12 US-11-056-470-2 Sequence 2, Appli     |
| c | 15    | 243   | 4.7   | 2352 12 US-11-052-554A-541 Sequence 541, App  |
| c | 16    | 240.5 | 4.7   | 215126 7 US-10-330-773-339 Sequence 339, App  |
| c | 17    | 240   | 4.7   | 3000 12 US-11-136-527-2740 Sequence 2740, Ap  |
| c | 18    | 238.5 | 4.6   | 2845 9 US-11-072-512-945 Sequence 945, App    |
| c | 19    | 235.5 | 4.6   | 15876 12 US-11-052-554A-560 Sequence 660, App |
| c | 20    | 235.5 | 4.6   | 200628 12 US-11-121-086-62 Sequence 62, Appl  |
| c | 21    | 234   | 4.6   | 3580 8 US-10-995-561-87 Sequence 87, Appl     |
| c | 22    | 233.5 | 4.5   | 173502 12 US-11-121-086-25 Sequence 25, Appl  |
| c | 23    | 232.5 | 4.5   | 2268 8 US-10-517-939-329 Sequence 329, App    |
| c | 24    | 232.5 | 4.5   | 63984 12 US-11-121-086-26 Sequence 26, Appl   |
| c | 25    | 230.5 | 4.5   | 3415 8 US-10-995-561-86 Sequence 86, Appl     |
| c | 26    | 230.5 | 4.5   | 126552 12 US-11-121-086-1 Sequence 1, Appli   |
| c | 27    | 230.5 | 4.5   | 191684 12 US-11-121-086-2 Sequence 2, Appli   |
| c | 28    | 227.5 | 4.4   | 5706 12 US-11-052-554A-519 Sequence 519, App  |
| c | 29    | 227   | 4.4   | 3921 12 US-11-052-554A-523 Sequence 523, App  |
| c | 30    | 227   | 4.4   | 4983 12 US-11-052-554A-521 Sequence 521, App  |
| c | 31    | 226.5 | 4.4   | 38920 7 US-10-330-773-228 Sequence 228, App   |
| c | 32    | 224   | 4.4   | 150468 12 US-11-112-908-56 Sequence 56, Appl  |
| c | 33    | 224   | 4.4   | 193789 12 US-11-112-908-55 Sequence 55, Appl  |
| c | 34    | 223.5 | 4.4   | 2304 12 US-11-052-554A-538 Sequence 538, App  |
| c | 35    | 223   | 4.3   | 116856 12 US-11-143-980-1 Sequence 1, Appli   |
| c | 36    | 221.5 | 4.3   | 2183 9 US-11-072-512-1064 Sequence 1064, Ap   |
| c | 37    | 221.5 | 4.3   | 2736 8 US-10-858-730-38 Sequence 38, Appl     |
| c | 38    | 221   | 4.3   | 168516 12 US-11-121-086-3 Sequence 3, Appli   |
| c | 39    | 219   | 4.3   | 1566 12 US-11-213-180-1 Sequence 1, Appli     |
| c | 40    | 219   | 4.3   | 3711 8 US-10-873-528-321 Sequence 321, App    |
| c | 41    | 218   | 4.2   | 11070 12 US-11-075-185-34 Sequence 34, Appl   |
| c | 42    | 218   | 4.2   | 78869 12 US-11-075-185-1 Sequence 1, Appli    |
| c | 43    | 217.5 | 4.2   | 11070 12 US-11-075-185-34 Sequence 34, Appl   |
| c | 44    | 217.5 | 4.2   | 78869 12 US-11-075-185-1 Sequence 1, Appli    |
| c | 45    | 217   | 4.2   | 26667 8 US-10-995-561-13375 Sequence 13375, A |

#### ALIGNMENTS

##### RESULT 1

US-11-052-554A-517  
; Sequence 517, Application US/11052554A  
; Publication No. US20050288866A  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatenIn version 3.3  
; SEQ ID NO 517  
; LENGTH: 9903  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-517

Alignment Scores:  
Pred. No.: 6.27e-05 Length: 9903  
Score: 282.50 Matches: 223  
Percent Similarity: 37.2% Conservative: 125  
Best Local Similarity: 23.8% Mismatches: 365



Query Match: 5.5% Indels: 226  
DB: 12 Gaps: 46

US-09-917-376-1 (1-957) x US-11-052-554A-517 (1-9903)

QY 56 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGlyGly----- 73  
DB 7399 AACCGAATCATCGCGCGCGCAACATCGCGGC-----TTTAACTCGGGATCGCA 7449

QY 74 -----AlaProGlyIleLeuTyValArgThrAspIleGly-----GlyMet 87  
DB 7450 AACACCGGTCCGGGCTAACCGCGGTGTCAACACATCGGTATCGGCAACCGGCAAC 7509

QY 88 TyrArgTTPAspAlaAsn-----GlyArgTTPileProLeuLeuAspTTPValGlyTTP 106  
DB 7510 TACAACATCGGTGTGCGCAACACCGGTAACTACAAC-----ATCGGCTTC 7554

QY 107 AsnAsnTTPGlyTyrAsnGlyVal-----ValSerIleAlaAlaAsp----- 120  
DB 7555 GGCAACACCGGCAACACACATCGGCATCGGCTGTCCGCGCACACACAGATCGGGTTC 7614

QY 121 ---ProIleAsnThrAsnLysValTTPAlaAlaValGlyMetTyrThrAsnSerTTPAsp 139  
DB 7615 GGCGCGCTGACCGCGC-----ATCGCCAAATGGGCTGTTC-----AACCTGGGCGAC 7665

QY 140 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTTPGlnIleThrPro 159  
DB 7666 AACAACTTTGGC----- 7677

QY 160 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 179  
DB 7678 -----ATGGCAACCGCGGCAACTTCAACACCGGCAATTCGC 7713

QY 180 AspProAsnAspAsnIleLeuTyrPheGlyAlaProSerTyrGlyLys-----GlyLeuTTP 198  
DB 7714 AACCGGCAACACACATCGGCTGTTCACACCGGCAACACACAGTCGGCATCTGG 7773

QY 199 ArgSerThrAspSerGlyAlaThrTTPSerGlnMetThrAsnPheProAspValGlyTTP 218  
DB 7774 CTGACCGCGCGCTGTTCGGCTTCAGTCCCTGAACTCCGGCGCGGCAACACCGGT 7833

QY 219 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTTP 238  
DB 7834 TTCTTCAACTCCGGCACCGGCAACACCGC----- 7863

QY 239 ValAlaPheAspLysSerSerSerLeuGlyGln-----AlaSerLysThrIlePheVal 257  
DB 7864 ---TTGTTCAACTCCGGCACCGGCAACACCGGCTTGTTCAACTCGGCGCACCGCAACGTC 7920

QY 258 GlyValAlaAspProAsnAsnProValPheTTPSerArgAspGlyGlyAlaThrTTPGln 277  
DB 7921 GGCATCGGCAACATGGGCACCGCGCTTCGGCTCGGCTATCCGCGGACAGCAGGTG 7980

QY 278 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 297  
DB 7981 GGCATCGGCGGCAACCACTCGGCGAGTTTC---AACATCGGCTTGTTTAACTCGGCGACC 8037

QY 298 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGly-As 317  
DB 8038 GGCAATGTCTGGCATCGGCAACTCGGCGACCGGCAACGTCGGCATCGGCAACCGGCGACC 8097

QY 317 pValTTPLeuPheSerValThrSerGlyThr----- 328  
DB 8098 GGCAACACCGGCAACCGGCAACACCGGCTGTTCACACCGGCGGCTG 8157

QY 328 p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs 342  
DB 8158 GTCAACACCGGCAATCGGCAACCGGCGCAACCAACACCGGCTGTTCACATCGGCGACC 8217

QY 342 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 362  
DB 8218 TTCAACACCGGCAATCGGCAACCGGCGCACTACACACCGGCTCTTACACACCGGCTAGC 8277

QY 362 aThrGlnIleSerTrp-----TrpProAspThrIleIle-----PheArgSerThr----- 377  
DB 8278 TACAACACCGGCAACACCGCGAGATACGCGGCGGCTTCATCATCGGCGACG 8337

QY 378 -----AspGlyGlyAlaThrTTPThrArgIleTTPAspTTPThrSer-- 391  
DB 8338 ATGAACACCGGCTTCTCTGGCGCGCACCGGCGAGGCTCTGGCGGCAACATACACC 8397

QY 392 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaG 406  
DB 8398 ATCACCATCGAGCGACCTCGCGGTCTCTCAATGTGCATCCCGTCAACATCCCATC 8457

QY 406 uProTTPLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTTP 426  
DB 8458 ACCG---CGGACATCACCAATGTCTCCATCCCGGCAATACGTTCGCCAGA-ATC----- 8508

QY 426 pMetAspLualaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 445  
DB 8509 ---GACCCACGCGAAGCGTGCATAGGCATCTCAGTGGCACCGTCTTGGCCCCGGT 8564

QY 445 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnI 464  
DB 8565 CGGTCCGATCATCTGCATCGGCGGACGCGTCCGCGCGCGTGGACACACCATCGAAT 8624

QY 464 eHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPr 484  
DB 8625 TGACTTCGGCCCC-----TCGCGCGCGATCAACCTCAACATCGGCA 8666

QY 484 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs 504  
DB 8667 GCCCGACGCTCCACCGGTGATCAACATCGTGGCGCGCGCGC----- 8709

QY 504 pValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVa 524  
DB 8710 -----GCCGCGCGGATC-----AGCAT 8726

QY 524 lAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe 544  
DB 8727 TCCGATCATCGACTTCGGCGCAGCG-----CCCGGCTTCTTCAACGCCACAC 8774

QY 544 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTTPPheGlnG 564  
DB 8775 CGGCGCG-----TCGTGGGCTTCTCCTCAACTCGGCGGTCTGG 8810

QY 564 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValaAla 578  
DB 8811 CAGCGCATCGGCTGTCTGAATTCGGCAACAACTCGGCGCTCTACAACTTCGCCACTAG 8870

QY 578 aSerAlaAspGlySerArgPheVal-----TTPAlaProG 590  
DB 8871 CAGCATGGGAATTCGGGCTTCCAAACTATGGGTGGTGGTGGTGGGCTGGCG----- 8925

QY 590 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 606  
DB 8926 -----AATTGGGCAACAGCATCTCGGCGCATCTA 8954

QY 606 pAlaAsnSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 626  
DB 8955 CAACACCGGCTGGGCGGCGGCAAAATGTC-----TCGGGCTTCTCAACATCGG 9005

QY 626 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyVa 644  
DB 9006 CACCAACTGGCTGGGTGGTGGAGACGCGCG----- 9039

QY 644 lThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPhe- 663  
DB 9040 -ACGAGACGACCTTCAGCTGGGCTTGGCCAACTCGGTTCT---GGAATCTGGGTAG 9095

QY 664 -----HisAlaValProGlyLysGlyAspLeuTTPLeuA 676  
DB 9096 CGCAACATCGGCAACTACAACTGGGCGGCGCAACATCGGCGTCTACAACTGGGCGAG 9155

QY 676 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlySerSerTTPSerA 693





6969 CAGCGAATAGACGGTTTATCTGCTGGACGTCTATCCGGTCCGATCGGCCCAACGTCGTA 7028 Db  
277 nAlaValProGlyAlaProThrGlyPheIleProHisLeuGlyGly---ValPheAspProVa 296 Qy  
7029 CTCGGTC-----GGCAGAAATCAACAGCCGGTGACCTGTCGTCGGCGG 7073 Db  
296 lAsnHisValLeuTyrlleAlaThrSerAsnThrGlyGlyProTyrrAspGlySerGl 316 Qy  
7074 CAATATCATCTCGGAATCCGACGCTCAGCACACCGCCGCGGTGATGCC----- 7125 Db  
316 yAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerProValProSe 336 Qy  
7126 -----GTCTTCATCGGCATCTCAGCGCTTCAAC-----ACCCCGC 7163 Db  
336 rThrAspThrAlaAsnAspTyrrPheGlyTyrrSerGlyLeuThrIleAspArgGln---Hi 355 Qy  
7164 GCTCAATAT-TCACAGATCCGTTGGTTTCCAGGTGCCGGCAGCATCGACGCGATCA 7222 Db  
355 sProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheAr 375 Qy  
7223 CCCT----- 7226 Db  
375 gSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrrProAsnAr 395 Qy  
7227 -----GTCCCCGGTGGTCTGACGTTCCCGGC-- 7253 Db  
395 gSerLeuArgTyrrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPr 415 Qy  
7254 -----GAATCGCTGCTGAACC-TGGATGTGTCTGTCGGCACCCCC 7293 Db  
415 oAsnProProValProSer-----ProLysLeuGlyTrpMetAspGluAlaMetAl 432 Qy  
7294 GCGGCCACCACTT-CCGCGATCACTGTTCCCGGAGATC----- 7329 Db  
432 alieAspProPheAsnSerAspArgMetLeuTyrr----- 443 Qy  
7330 -----CCGCGGAACCCGACGCGCAACTTACGTATCCGCGGACATCCCGCTGAT 7382 Db  
443 ----- 443 Qy  
7383 CAACATCCCGCCACCCCGGCGATTGGGAACACCACCACCGCTCCGTGCGGGCTTCTT 7442 Db  
444 -----GlyThrGlyAlaThrLeuTyrrAlaThrAsnAspLeuThrLy 457 Qy  
7443 CAACACCGGGCGGGCGGCTCGGGTTTCGGCAACTTCGGCGCAACATGTCGGGGTG 7502 Db  
457 sTrpAsp-----SerGly-----GlyGl 463 Qy  
7503 GTGGAAACGAGCGCACACCGCTTTGGCAGCGCGGGTTCGGGTATTGCCAACGTCGGCAC 7562 Db  
463 nIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSe 483 Qy  
7563 ACTGCATCGGGTGTCAACCTCGTTCGGGGCTGTCGGGGATCTTACAAACACCAGCAC 7622 Db  
483 rProProSerGlyAlaPro---LeuIleSerAlaLeuGlyAspLeuGly----- 498 Qy  
7623 GCTGCGCTTGGAGCGCGGGTGTGTGTCGGGCTTGGCAACGTCGGTGATCACTGTC 7682 Db  
499 -GlyPheThrHisAlaAspValThrAlaValProSerThrIlePhe-ThrSerProValP 518 Qy  
7683 GGGCTTGTGGCTTCCAACTGGGGGAAAAACCCCATCACCATCGTCAACATCGGGTTGGC 7742 Db  
518 heThrThrGlyThrSerValAspTyrrAlaGluLeuAsnProSerIleIleValArgAlaG 538 Qy  
7743 TAACGTCGGCAACGGCAACCTCGGCTTCGGCAACATCGGC-AACCTCAACCTGGGTGCGG 7801 Db  
538 lySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyG 558 Qy  
7802 CCACATT-----GGCAGCGTGAACCTGGGATTCGGCAACATTGGCG 7843 Db  
558 lyLysAsn-TrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrValAla 577 Qy  
7844 ACGTGAACCTGGGTTCGGCAACATCG---GCGCGGCAACGTCGGGTTCGGCAATATCG 7900 Db

578 Ala-----SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGln 594 Qy  
7901 GCGATGCCAACTTCGGGTTCGGAATTCGG---GTCTGGCGCGGCTTGGCGCGCATGG 7957 Db  
595 ProVal-----ValTyrrAlaValGlyPheGlyAsnSerTrpAla--- 607 Qy  
7958 GCAATATCGGGCTGGGCAACCCCGCAGCGCAACGTCGGTGGCCCAACATGGGCCCTGG 8017 Db  
608 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 627 Qy  
8018 GCAACATCGGGTTGGCAACACCGCACCACTCGGATCGGGTTCACCGCGGACA 8077 Db  
628 PheTyrrAlaLeuSerAsnGlyThrPheTyrrArgSerThrAspGlyValThrPheGln 647 Qy  
8078 -----ACCATCGC 8086 Db  
648 ProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValPro 667 Qy  
8087 GCATCGCGGGTTGAACCTCCG-----GCATCGCAACATTTGGCCCTGTTCACACTCG 8137 Db  
668 GlyLysGluGlyAspLeuTrpLeuAlaAspSerSerGlyLeuTyrrHisSerThrAsnGly 687 Qy  
8138 GCACAGCA-----ATATCGGCTTCTTCACTCCGGGACTGCGCAACTTCGGGT 8185 Db  
688 GlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLys 707 Qy  
8186 TGTTCAACTCCGCGAGCTTACAAACCCGATCGCAACTCCG-----GGTGG 8233 Db  
708 SerAlaProGlySerSerTyrrProAlaValPheValValGlyThrIleGlyGly----- 725 Qy  
8234 CCAGCACCGGGTTGGTCAACCGCGCGGC-TTCAACACCGGTGTGGCAAAACCGCGGTG 8292 Db  
726 ---ValThrGlyAlaTyrrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAsp 744 Qy  
8293 TACAACACGGGACGTTCAATGCTGGTGACACCAACACCGG-TGGCTTCAACCCAGGCGAG 8351 Db  
745 AspGlnHisGln-TyrrGlyAsnTrpGlyGlnAlaIleThrGly-----AspHi 760 Qy  
8352 CACCAACACCGGCTGGTTCAACACCGGTAAACCCCAACACCGGCTGCGCAACCGCGGCA 8411 Db  
760 sAlaAsnLeuArgValTyrrIleGlyThrAsnGlyArgGlyIleVal----- 776 Qy  
8412 TGTCAACACCGCGCCCTCATCACGGGCAACTTTAGCAACGCGCATCTTATGCGGGGCAA 8471 Db  
777 -TyrrGlyAspIleGlyGly-AlaProSerGlySerPro-----SerProSerValS 793 Qy  
8472 TTACGAGGGCTTGGCGGCTTCTCTTCCGGTACCCCATTCCTCGCTGTTCGCGCGGTGG 8531 Db  
793 erProSerAlaSerProSerLeuSerProSerProSer-----ProSerSerSer---P 810 Qy  
8532 CGCCGAGTCAACCGCGACATCGGCCCGCCACCATCATTCGCCCCATCCATCCCGCTC 8591 Db  
810 roSerProSerProSerProSerProSerProSer-----SerSerP 824 Qy  
8592 CATCCCGTTGGGCTTCGCGCGATCGGCCACATCGGGCCGATCAGCATCCCGCAACATCGC 8651 Db  
824 roSerPro-----S 827 Qy  
8652 CATCCCTCGATCCACCTGGGATCATGCCACCTTCGACGTCGCGCCCTATCACCGTGA 8711 Db  
827 erProSerProSerProSer-----ProSerArgSerProSer- 839 Qy  
8712 CCCCATCCCTTCACTCCCTGCGCTTAAGTTTGGATGCTGCCGCTCTCGGAGATCAGGAT 8771 Db  
840 -----ProSerAlaSerProSerProSerProSerProSerProSerProSerS 858 Qy  
8772 GACGTCGGAGAGAGCTCCGGATTCAGGTCAAGTCAGCCAGCTTTTCATCTTCGCGGTGG 8831 Db  
858 erSerProSerProThrProSerSerProValSerGlyValLysValGlnTyrrL 878 Qy  
8832 AC-----CCGACGCGCATGTCGCG-----GGGCGGAGGTCTCCATCT 8867 Db

```

Qy 878 yAsnAsnAspSerAlapro-GlyAspAsnGlnIleLysProGly-LeuGln----- 894
Db 8868 TCACCACT-----TCACCGTGGGACCACTCACTGAACCGGACAGCACTGCATCTCCC 8921
Qy 895 -----ValValAsnThrGlySerSerValAspLeuSerThrValThrValArgTy 912
Db 8922 CGGATTACCACTCCACCGGACCATCCACATCGGCTGGCGTGTGCTGACCATCC 8981
Qy 912 rTrpPheThrArgAspGlyGlySerSerThrLeuVal-TyrAsnCysAspTrpAlaAlaI 932
Db 8982 GGGCTTCACCATCCCGGGCGGC-----ACCGTGAATCCCACTCCCGCTGGGCTCGG 9035
Qy 932 le-----GlyCysGlyAsnIleA 938
Db 9036 TTGTTCGGGGCGGACCCACCCCTTTGATCTCCGACGGTGTATCGACCGGATCC---C 9092
Qy 938 rGAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThrTyr 954
Db 9093 GGTGGAGTTACCGCCAGCACCATCGGCCCGCTGACGCTCCCGATTT 9142

RESULT 3
US-11-052-554A-525
; Sequence 525, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 525
; LENGTH: 11151
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-525

Alignment Scores:
Pred. No.: 0.000201 Length: 11151
Score: 273.00 Matches: 279
Percent Similarity: 35.2% Conservative: 115
Best Local Similarity: 24.9% Mismatches: 385
Query Match: 5.3% Indels: 349
DB: 12 Gaps: 65

US-09-917-376-1 (1-957) x US-11-052-554A-525 (1-11151)
Qy 79 TyrValArg---ThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrp 97
Db 2399 TGGGTGGCGGCAACATCGGTGGGCGACACCTGGGCTGG---GCAATCTCGGG----- 2449
Qy 98 IleProLeuLeuAspTrpValGlyTrp-----AsnAsnTrpGlyTyr--- 111
Db 2450 -----ACGGCAACCTGG---GGTGGGCAACATCGGCGCATCGGCAACTGGGGTTTGCCA 2500
Qy 112 -----AsnGlyValValSerIleAlaAlaAspProIle 122
Db 2501 ACTTGGGCTTGACGCGCGCGGGGTGGGCAATGTTGGTGGCAATGCGGCA 2560
Qy 123 AsnThrAsnLys-----ValTrpAlaAlaValGlyMetTyrThr----- 135
Db 2561 TCACACACTATGCTTGGCGAACAATGGGTGGGCAATATGTTGGTTGGCCACACCGGCA 2620
Qy 136 -----AsnSerTrpAsp-----ProAsnAspGlyAla---Ile 145
Db 2621 CGGGCAACATCGGGATCGGGCTGTGTCGGGGACCATCGGACCGGGATCGGGGCTTGAAC 2680

```

```

Qy 146 LeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGly 165
Db 2691 CGGGCATCGGCAATATCGGGTTGTTCA-----ACTCCGGCACCGGCAACGTCCGGT 2731
Qy 166 GlyAsnMetProGly-----ArgGlyMetGlyGluArgLeuAlaValAspProAsn 182
Db 2732 TCTTCAATTCGGGACCGGCAACTTCGGCATCGGAACTCCGGCGCTTCAACACCGGGA 2791
Qy 183 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAsp 202
Db 2792 TCGGT-AAT-----AGCGAAACGGCCAGCACCGGGCTCTTC---AATGCGGC 2835
Qy 203 SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn 222
Db 2836 AGCTTCAGCAGCGGATCGGCAACACACTGTGACTACAAACACACGCGGAGCTTCAACGCGGC 2895
Qy 223 ProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAsp 242
Db 2896 GACACCAACACCGGTGGCTTCAACCCGGGCGGCATCAACACCGGCTGG-----TTCAAC 2949
Qy 243 LysSerSerSerLeuGlyGlnAlaSerLysThrIlePhe-----ValGly 258
Db 2950 ACCGGGATGCAACACACCGGGTTGGCCAAACGGGCGGACCTTCGGACACCGGCGCTTCATG 3009
Qy 259 ValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly----- 273
Db 3010 ACGGCGACTACAGCAACCGGCTGTGTGGCG-----GGCGGTACGAGGCGCTGGTC 3063
Qy 274 -----AlaThrTrpGlnAlaValProGly 281
Db 3064 GCGGTCCGCGTGGGCGCCACGATCTCCCAATTCGCGGTACCGTGCACGATCGGCGGG 3123
Qy 282 ---AlaProThrGlyPheIleProHisGlyValPheAspProValAsnHisValLeu 300
Db 3124 GTGGCGCGCTGCATGTGGCGGCC-----GTCCCGGTATCCCGCCCGTGCACGTC--- 3171
Qy 301 TyrIleAlaThrSerAsnThrGly---GlyProTyrAspGlySerSerGlyAspValTrp 319
Db 3172 GAGATCACCGACGCGCCACCGTCCGCTGGTGGTTCGTTCCCGTCCACCGATCAGCATCCC 3231
Qy 320 LysPheSerValThrSerGlyThrTrpThr-----ArgIleSerPro 333
Db 3232 TCACCTCCCATCGGCAGCATCACCGGAAGCGTGGACCTGGCGGCAACACCATCTCGCG 3291
Qy 334 ValProSerThrAspThrAlaAsnAspTyrPheGly----- 345
Db 3292 ATTCGCGCTCTTGACCCCGCTCGCGGTTCGATAGGGCTTTTTCGAGCGCTTCGCGCTC 3351
Qy 345 ----- 345
Db 3352 AGTGACCACTTATCACCATTGATCGGTTCCAAGTTGTTGCGGTGCTTGTTCCTAGAG 3411
Qy 346 -----TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAla 362
Db 3412 AACATCATTTGGCGCGCTCACGGTTAGCGGTAG-----ATATTGGTCAACCCCG 3462
Qy 363 ThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThr 382
Db 3463 ACACCAATT-----CCCTAACCTCAACTTGGACACCCAC----- 3498
Qy 383 TrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAsp 402
Db 3499 -----CGTGACCGCTTTTCCCGAATGGT-----TTCACC 3528
Qy 403 IleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProValProSerPro 422
Db 3529 ATTCGCGGCAACACCCCGTACGGTGGTATGAGG---TCGCCAACGAGGGTTCACT 3586
Qy 423 LysLeu-GlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLe 442
Db 3587 TCTTCCGGGTGGGTGACCTTTTCGCGGGCTCCGC-----CGGGGTCAACGGACTGT 3640
Qy 442 uTyrGlyThrGlyAlaThr-----LeuTyrAlaThrAs 453

```

Db 3641 CCGTGGGGCTGACGGGTTACACCTGTTGCCAGCGGTTACCTCGACACCGTGGCCG 3700  
Qy 453 nAspLeuThrLysTrpAsp-----SerG1 461  
Db 3701 CGACCTTCGACGGCACCATCTCATCGGGGATATCCGATCCCGATCATCGATGTGCCG 3760  
Qy 461 yGly-----GlnIleHisIleAlaProMetValLysGlyLeuGluThrAl 477  
Db 3761 CGGTGCGGGGGTTCGGCAACACACCGGCCCTCATCGTCCGGGTTCTTCAACACCGGCG 3820  
Qy 477 aValasn-----AspLeuIleSerProPr 485  
Db 3821 GCGCGGTGGATCGGGTTCGCCAACGTCGCGCGCGCACGTCGGGCTGGTGAACCA-- 3878  
Qy 485 oSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly-----PheThrHisAl 503  
Db 3879 ----GGGGACGACGTGTAGCAGG---GGCGGGCTCGGGAGTTGCCAATGCCGGCACGC 3931  
Qy 503 aAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrse 523  
Db 3932 TGAGCTCGG-GCGTGTGAACGTGCGGCTCGGGGATCTCCGGGTGGTACACACCGACACC 3990  
Qy 523 rValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProse 543  
Db 3991 CTGGGAGCGGCA-----CCCCGGCGGTGCTCTCGGGCATCGGCACACCTCGGCCAG 4041  
Qy 543 xSer-----GlnProAsnAspArgHisValAlaPheSerThrAs 556  
Db 4042 CAGCTGTCCGGGTTCTTGCAATGGACCGTCTCAACCGGAGCCCATGTTCATATC 4101  
Qy 556 pGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrVa 576  
Db 4102 GGGTGGCGCATGTGGCGGT---TCAACACCGGGTGGGCAATGGGGACCTCAAC 4158  
Qy 576 lAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProVa 596  
Db 4159 TGGGGTGGCGCAACATCGCGCGGAGAACCTGGGCTG-GGCAATCTCGGACGCGG-- 4215  
Qy 596 lValTyrAlaValGlyPheGlyAsn-SerTrpAlaAlaSerGlnGlyValProAlaAsnA 616  
Db 4216 -----AACGTGGGTTCGGCAACATCGGTGCGGCAACGTGGGTTCGCCAACT-- 4264  
Qy 616 laGlnIleArgSerAspArgValasn----- 624  
Db 4265 -----CGGTCGGCGTGGGCTGGCGGCTGGGCAACGTGGGTT 4307  
Qy 625 -----ProLysThrPheTyrAlaLeuSerAsn- 633  
Db 4308 GAGCAATGCGGCGAGCAACACTGGGGGCTGGCCAACTGGGTGTGGGCAACATCGGTT 4367  
Qy 634 --GlyThrPheTyrArgSerThrAspGly-----GlyValThrPheGlnProValA 650  
Db 4368 GGCCACACCGGACGCGGCAACATCGGGATCGGGCTGGTGGCGGATACACAGACGGCAT 4427  
Qy 650 laAlaGlyLeuProSer---SerGlyAlaValGlyValMetPheHisAlaValProGlyL 669  
Db 4428 CGGC-GGCTCAACTCGGTAGTGGCAATATCGA---TTGTTCAATTCGGGACCGGCA 4483  
Qy 669 yeGluGlyAspLeuTrpLeuAlaAlaSerSer---GlyLeuTyrHisSer----- 684  
Db 4484 ATGTCCGGTTCCTCAACACCGGACCGGCACTTCGGACTGTTCAACTCGGTAGTTTCA 4543  
Qy 685 --ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValG 704  
Db 4544 ACACCGGATCGGTATAGCGGAACCGGACGTACTGGGCTCTTCAATGCCGCAATTTCA 4603  
Qy 704 lyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleG 724  
Db 4604 ACACCGGATCGGCCAACCCCGGTCGTACACACCGGCGAGCTTCAATGTCCGTGATACCA 4663  
Qy 724 lyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnA 744

Db 4664 AC-----ACCGGTGGTTTCAACCGGCGGACATCAACACCGGCTGG-----TTCAACA 4711  
Qy 744 sAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuA 764  
Db 4712 CCGG-CATTATGATACGGG-----CACCCGCAACACC 4743  
Qy 764 rArgValTyrIle-GlyThrAsnGlyArgGlyIleValTyr---GlyAsp----- 779  
Db 4744 GCGCCCTCATGTCCGGGACCGACAGCAACGCGATGCTGTGGCGGCGCACACGAGGGC 4803  
Qy 780 lIleGlyGly-AlaProSerGlySerProSerValSerProSerValSerProSerProse 799  
Db 4804 CTGTTCCGCTGTCTATGCGATCACGATCCCGCAAT---TCCGATCCGATCACACG 4860  
Qy 799 rLeuSerProSer---ProSerProSerSerPro----- 810  
Db 4861 ACTGGCGTATCGGCCCATCGTATCCGGACACACGATCTTCCGCGGCTGCACCTG 4920  
Qy 811 -----SerProSerProSerProSerProSerProse 821  
Db 4921 CAGATCACCGGCGAGCGGACTACAGCTTCAACGTCGCGGACATCCCATCCCGGCATC 4980  
Qy 821 rSerSerProSerProSerProSerProSerProSerProSerProSerProse 841  
Db 4981 CACATCGGCATCAATGGCGTCGTACCGTCCGCTTACCGCCCGGAGGCCA---CCCTG 5037  
Qy 841 rAlaSerPro-----SerProSerProSerProSerProSerProse 855  
Db 5038 CTGTCCGCCCTGAAGATACGCTAGTTCATGAGTTCGGGCCCATCAG-CTCTCGAA 5096  
Qy 855 rProSerSerProSerProSerProSerProSerProSerProSerProSerProse 875  
Db 5097 TATCGATATTCGCGCCCATGGATTTCAAGTTAGGCTGCGGCTTCTTGTCTTATCACGG 5156  
Qy 875 lGlnTyrLysAsnAsnAspSer-----AlaProG1 885  
Db 5157 CCAACT-CGGACCAATTCATCTTGAGCCAAATCGTGGTGGCGGATCGGTGTGCCCTGG 5215  
Qy 885 yAspAsnGlnIleLysProGlyLeuGlnVal----- 896  
Db 5216 AGATCGAGCCCATCCCTGGATCGGATTCGTTGAGTGAGTCGATTCCTATCCGCAATAC 5275  
Qy 897 -----AsnThrGlySer----- 900  
Db 5276 CTGTTGATATTCGCGCTCGGTATCGATGGGATTCATGTCGGAAGTGTGCGGATCG 5335  
Qy 901 --SerSerValAspLeuSerThrValThrValA-gTyrTrpPheThrArgAspGlyGlys 920  
Db 5336 ATGCGTCGTTGGACATCCGGCGGTACGATC-----A 5368  
Qy 920 eSerThrLeuValTyrAsnCysAspTrpAlaAlaIleGlyCysGly---AsnIleArgA 939  
Db 5369 CAGCACCAACCATT-----TCCGCGATCCCGTGGGCTTCGACATTCGCA 5413  
Qy 939 laSerPheGlySerValasn-----ProAlaThrPro 949  
Db 5414 CCAGTGGCGGACCCCTCAACATCCCGATCGACATCCCGCGGCGCGCG 5463

RESULT 4  
US-11-052-554A-518  
; Sequence 518, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 518

; LENGTH: 6615

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis H37Rv

US-11-052-554A-518

## Alignment Scores:

Pred. No.: 0.000183 Length: 6615  
 Score: 270.50 Matches: 263  
 Percent Similarity: 32.4% Conservative: 105  
 Best Local Similarity: 23.1% Mismatches: 363  
 Query Match: 5.3% Indels: 412  
 DB: 12 Gaps: 56

US-09-917-376-1 (1-957) x US-11-052-554A-518 (1-6615)

Qy 53 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValaspGlyIleValPheAsnGlu 72  
 Db 5 ACCTGGTCTCC-ACAACGTCGGGAATGTCGGGCTTCCTC-----AACGTC 48  
 Qy 73 GlyAla-----ProGlyIleLeuTy-ValArgThrAspIleGlyMetTyArg-Tr 90  
 Db 49 GCGCGCTGGATCGGGTGTGGCGAATGTGGCAACCACTCTCGGGTATCTACAACGTG 108  
 Qy 90 pAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpG1 110  
 Db 109 GGCACGTCGGACCTCTCGACCGCCGCGTTAACTCCG-----GGTGGCAA- 154  
 Qy 110 yTyAsnGlyValSerIleAlaAlaAspProIle-----AsnThrAs 125  
 Db 155 -----ATATCGGAACCAATATTGTCGGCCTGCTCGCGACGCGC 192  
 Qy 125 nLysVal-----TrpAlaAlaValGlyMetTyTrpThrAsnSerTrpAspPr 140  
 Db 193 GCGGGTACTCGGGTATTAACTTGGGTGTGGCAACCAACCGCAACCTCAACGTGGCTTC 252  
 Qy 140 oAsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThrProLe 160  
 Db 253 GCAAGTCTCGCGGCTTTAACTTCG----- 277  
 Qy 160 uProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValas 180  
 Db 278 -----GGGCGCC 285  
 Qy 180 pProAsnAsnAspAsnIleLeuTyPheGlyAlaPro---SerGlyLysGlyLeuTrpAr 199  
 Db 286 ACCATCGCCACACAAACGTCGGCATCGGGAACACCGGAATCTTCGATGTCGGCTGGCG 345  
 Qy 199 gSer-----ThrAspSerGlyAlaThrTrpSerGlnMe 210  
 Db 346 AACCTGGCGAGCTACAACATCGGCTTCGGAACCTTCGGCGACGACCACTGGGCTTCGGC 405  
 Qy 210 tThrAsnPheProaspValGlyThrTyIleAlaAsnProThrAspThrThrGlyTyGr1 230  
 Db 406 AACTCG-----CCAGCTACAACATCGGCTTCGGCAACGTCGCAACG- 448  
 Qy 230 nSerAspIleGlnGlyValTrpValAlaPheAspLysSer----- 245  
 Db 449 -----ACAATCTGGTTCGCTAACCGCGCGCGCAACATCGGCTTT 492  
 Qy 246 -----SerSerLeuGlyGlnAlaSerLysThrIlePheValG1 258  
 Db 493 GCGAACACCGCGACGAACAATGTGGTGTGGGAACACCGGCGCAACAATGTCGGCATC 552  
 Qy 258 yValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAla----- 274  
 Db 553 GGGCTCAGCGCAACGGAC-----AGATCGGGTTCGGCAGCTTCAACTCG 597  
 Qy 275 -----ThrTrpGlnAlaValPr 280  
 Db 598 GGCAGCGGAACATCGGCTGTTCAACTCGGGCAGCAACACATCGGATTTCTTCAATTC 657

Qy 280 oGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValle 300  
 Db 658 GGCAGCGCAACTTCGGC-ATCGAAACTCGGGCAGCTTCAAC----- 699  
 Qy 300 uTyIleAlaThrSerAsnThrGlyGlyProTyArgGly-----SerSerGlyAspVa 318  
 Db 700 -----ACTGGCATCGGAACACCGGCAACACCAATACCGGCTTATTCACTCGGCGAC- 753  
 Qy 318 lTrpLysPheSerValThrSerGlyThrTrp-ThrArgIle-----SerProValProS 336  
 Db 754 -----GTCACACCGGGCGCTTCAACCCGGGCGAGCTTCAACACCGGTAGCTT 800  
 Qy 336 eThrAspThrAlaAsnAspTyPheGlyTySerGlyLeuThrIleAspArgGlnHisP 356  
 Db 801 CAACACCGGCGACTTCAACACCGGTGGCT--TCAATCGGGCAATACCAACACCGGCTA 857  
 Qy 356 roAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIle-----P 374  
 Db 858 CCTCAACATTTGGCAACTACAACACCGGCATCGCCAACACCGCGAGCTTGAACACCGGGC 917  
 Qy 374 heArgSerThrAspGlyGlyAlaThrTrp-----ThrArgIleTIPa 388  
 Db 918 TTTCAATACCGGAACACTACAGCAACGGTGTCTTAAAGCGCGGATTACCAGGCGCTGG- 976  
 Qy 388 sPTpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAlaGluProT 408  
 Db 977 -----TCGGCTCAACCT 989  
 Qy 408 rpLeuThrPheGlyValGlnPro-----A 416  
 Db 990 GGTGATCATGATGCTCTCCCATTAAGCTCTCGGCGTGAATATTCCCATCGATATCCCGAT 1049  
 Qy 416 snProProValProSerProLysLeuGlyTrp----- 426  
 Db 1050 CACCGCTCGCGCGCAACATCACCCTTATGGCGTACGATTCCGCCACCGCGCGATAT 1109  
 Qy 427 -----MetAspGluAlaMetAlaIleAspProPheAsnSerA 439  
 Db 1110 CGTCCTTTCTCAATAGCGGCGCACGAGCCACCTTTGGCCCATACCATTTCCGACAT 1169  
 Qy 439 spArgMetLeuTy- GlyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTr 458  
 Db 1170 CACGGTTCTCGGCCCCACGAGCAGCAGTCGCCATAGGAGGGCGCAATACCGCGATCACC 1229  
 Qy 458 pAspSerGlyGlnIleHille-----AlaProMetValLysGlyLeuG1 474  
 Db 1230 AACTGGCGGTGGCGCATTAGGATCCCGCTCATCATGATATCCCGCGGCGCGAGTTTCGG 1289  
 Qy 474 uGluThrAlaValAan----- 479  
 Db 1290 AAACCTCGAACCAACCCCTGCTCAGTTTCTTCAATACCGGCGCGCGCGCTTCGGG 1349  
 Qy 480 -----AspLeuIleSerProSerG1 487  
 Db 1350 CTTTCGGCAACTTCGGCGGCGCAATTCGGGCTTTTGGAACTTGGCTTCGCGACCTCGG 1409  
 Qy 487 yAlaProLeuIleSerAlaLeuGlyAspLeuGly----- 498  
 Db 1410 GCGCTCGGGGCTCTCAACGTCGGCGCTGGATCAGGTCTGGCGAACCTGGGCAACCCAC 1469  
 Qy 499 -----GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerPr 516  
 Db 1470 CGTCTCGGGGTTCTACAACACCGACGACGTCGGACCTCGCGACGCGCGCTTCAATTGAG 1529  
 Qy 516 oValPheThrThrGlyThrSerVal----- 524  
 Db 1530 CCTGGCGCAATCAGCACCATGATCCCGGCTGCTGCGGACGACGACGCGGACCATGTT 1589  
 Qy 525 -----AspTyAlaGluLeuAsnProSerIleIleValArgAlaG1 538  
 Db 1590 CCTCAACCTGGGCTTGGCAACACCGACCTTCAACCTCGGCGCAT----- 1635



80 TGGTGGCCG---ACGGCGGAGCGCATCGCGGGTTCCTCCGGAGGACCGGGGCTGGGACC 136  
387 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 406  
137 TGGCGGGGCTCTCGACTCCGACCGCGAGC----- 166  
407 ProTrpLeuThrPheGlyValGlnProAsnPro-----ProValProSerPro 422  
167 CCGTGG-----GCAAGTCCTATGTGCGGAGGGCGGTTTCCTCACCG 208  
423 LysLeuGlyTyrMetAsp-----GluAlaMetAlaIle 433  
209 AC-GGGGGCGGATTGACCGCGCATCTTCGGCATCTCGCCCCGTGAGCGCTGCGGATG 267  
434 AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsn 453  
268 GACCCG-----CAGCAGCGGTGCTGCTGGAGACCGCG----- 300  
454 AspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaPro---MetValLysGly 472  
301 -----TGGGAGACCTTCGAGAATCGCGGAATCGACCCCGGGTTCGTCGACCGC 348  
473 LeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSer 492  
349 ACCGACGTGCGTGTTCAGCGGAGTGATGTACCAAGATTACGGGGCCGACCGCGGAGC 408  
493 AlaLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSerThrIle 512  
409 GCGGGGAGGCGCTGGAGGG-----CATCTGGCGTGGGACGCGCGGGAGGTGTC 462  
513 -----PheThrSerProValPheThrGlyThr--- 522  
463 TCCGGGGGGTGGCTACGCGCTGGCGCTGACCGGGCCCGCGGTGACCGTGACACCGCC 522  
523 TGCTGCTCTCCCTGGTAGCGCTGCACCTGCGGTTACAGCGGTGTCGCGACGGCGAATGC 582  
533 IleIleValArgAlaGlySerPheAspProSerGlnProAsn----- 547  
583 TCGTGGCGCTCGCGGGGTGTCGGGTGATGAGCAGCGCGCTGCAAGTCTTCGCGGAGGCGCCAC 702  
548 AspArgHisValAlaPheSerThrAspGly----- 557  
643 TCCGCGCAGCGTGCGCTCGCCCCCGACGCGCGCTGCAAGTCTTCGCGGAGGCGCCAC 702  
558 GlyLysAsnTrpPheGln----- 563  
703 GGCACCAACTGCTCGAGGGGTGTCGGGTGCTGCTGGAGCGGCTGTCGATGCCCGC 762  
564 -----GlySerGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAsp 581  
763 CGCAATGGGCATGAGGTGCTCGCGCTGCTCGTGGC-----ACGCGCTGAACAGGAC 816  
582 Gly---SerArgPheValTrpAlaProGlyAspProGlyGln----- 594  
817 GCGCCAGCAACGCGCTGACCGGCCCAACGCGCGCTCCAGGAACGGGTGATCGGCAG 876  
595 -----ProValValTyrAlaValGlyPhe 602  
877 GCGTGGCGAACCGCGGCTGACGCTGGCCGATGTGGACGCGGTGAGAGCCACCGGCACG 936  
603 GlyAsnSer----- 605  
937 GGCAGAGTCTCGGCGACCGGATCGAGGCCAGGCACTCTTGGCCACCTACGGGCGAGGAG 996  
606 -----Trp----- 606  
997 CGGCGGAGGATCAGCGCTGCTGGGGTCTGGTGAAGTCGAACATCGGGCATGGCGAG 1056  
607 -----AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsn 624  
1057 GCGGGGCGGCGCGCGGTGTCATCAAGATGTCTCAGGCCATCGCGGCGGCTACTG 1116

625 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 644  
1117 CCAAAACCTCCAGCGCCAGCAGCCACGACCAAGTCTGACTCGACTGAGTGGCGGTG 1176  
645 -----ThrPheGlnProValAlaAlaGlyLeu 653  
1177 TCGTACTGTTCGAGGCGCGGCTCGCGGAGACGGGACACCCCGCGCGCGGAATC 1236  
654 ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeu 673  
1237 TCCTCTTCGGCGTACGCGGAGCAAGACGACAGTGGT-CCTGGAACAGC----- 1286  
674 TrpLeuAlaAlaSerSerGlyLeuTyrHisSerThr-----AsnGlyGlySerSerTrp 691  
1287 -----ACCCCTGGAAGCGCTGCACCCGAAACACAGCGGAGCAGCGCGGCGCTCTCG 1340  
692 SerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGly 711  
1341 GCTCGTGGCCACGCGGCGGTAGTCCGTC----- 1370  
712 SerSerTyrProAlaValPheValGlyThrIleGlyValThrGlyAlaTyrArg 731  
1371 -----GGTCTGTCCGC 1382  
732 SerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGln-His---GlnTyrGln 750  
1383 CAAAGACTCTCGCGCGCTCGCGCTCAGCAGAGCGTCTGGTCAGCCATCTGGAGTCCGG 1442  
750 yAsn-----TrpGlyGlnAlaIleThrGlyAspHisAlaAs 762  
1443 GAGCAGACCAACCGGTCGATGTGGCTGGTGGTGGCCACACCGCGGCGGTGGA 1502  
762 nLeuArgValTyrIleGlyThrAsnGlyArgGly----- 774  
1503 GCACCGCGCGGTCTCTGGCGCAGGATCGGAAGGAGGATGGCGACGCGCGGCGTCT 1562  
775 -----IleValTyrGlyAspIleGly----- 781  
1563 GCGGAGGGGCGGCTGACCCGCTCTGTGTACCGGACAGACGGAACAGACGCGCAAAAC 1622  
782 -----GlyAlaPro-SerGlyS 787  
1623 CGTGTTCATCTTCCCGCGCAAGCGCCCAATGGTGGGATGGAGGCCCACTCTCTCAA 1682  
787 exProSerProSerValSerPro----- 794  
1683 CACCTCACCGCTTCTCGCGCGCGCTGCTGAGTGGCGGATGCTCTAGCGCGGTATAC 1742  
795 -----SerAlaAs 797  
1743 CGACTGTCGTCATCGAGCTCATCCCGGACGCCCGAGCTCCCTCGCTTGGCGGTGT 1802  
797 exProSerLeuSerProSerProSerProSerProSerProSerProSerProSerPro 812  
1803 CGACTGTCAGACGCGGACCTTGGCGGTGCTGCTCTCCCTCGCGCATCTTGGCAATC 1862  
813 -----SerProSerProSerSer----- 818  
1863 CGTGGGATCCACCCCGAGCGCTCATCGCCACTCCCAAGGCGAAATCGCCGCGCGCTG 1922  
819 --SerProSerSerProSerProSerProSerProSerProSerProSerProSerProSerPro 832  
1923 CGTCCCGGACACCTTCACCTCACCACCGCCGCAAAATCGTCACCTCTCGCAGCCAGAC 1982  
832 roSerProSerArgSer-----ProSerP 840  
1983 CATGCCACCACTCGCGGACGCGGCGATGATGTCTCGCCACCCCGCGGACAC 2042  
840 roSerAlaSerPro----- 844  
2043 CATCGACCTCACCACTGGCAGCGGCAAACTCTGATCGCGGCACACACAGCCCAACGC 2102





```
Db 5599 GCGCGATGGTCTGGGTGGCTGAATCAGCTCGCTCGGCTGCTGGGTGGGGTTCG 5658
Qy 448 ThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisAla 467
Db 5659 GGTGTTTCAACGCGCGGCACGCTG-----CATTCGGGTGTGCTGAATTCGCG 5706
Qy 468 ProMetValLysGlyLeuGluGluThAlaValAsnAspLeuLysSerProSerGly 487
Db 5707 TCTGCGATGTCGGGGCTGTTCAACACCGAGCGTGTGGGGTTG-----GGT 5751
Qy 488 AlaPro--LeuLysSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr 506
Db 5752 GCGCGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 5790
Qy 507 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 526
Db 5791 -----CAGTTGTCGGGATGTGGCGAGCGGAGCGCG----- 5823
Qy 527 AlaGluLeuAsnProSerIleValAlaGlySerPheAspProSerSerGlnPro 546
Db 5824 -----CTGCATCAGGTGTGGTCTCAATTCGGGTGGCGGATGTGGGTG----- 5871
Qy 547 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnThrPheGlnGlySerGlu 566
Db 5872 -----GCAATGTCTGGGTGGGCAATGTCTGGGAGCTTTAACTCTGGGTGGCGGCAACGTC 5925
Qy 567 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 586
Db 5926 GGTGGTTCATATGGTGGTGGCGGCAATATCGCGGCAACATGCGGGTTCGGCAATGTC 5985
Qy 587 ---TrpAlaPro-----GlyAsp-----ProGlyGlnProValValTyr 598
Db 5986 GGTGGGCGCAATTTGGCTCGGGAATTCGGGGTTAAGCGCGGTCTCATGGGTTCGGT 6045
Qy 599 AlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIle 618
Db 6046 AATATCGGTGTGGTAATCGCGGAGCTACAAATTCGGGTG-----GCCAATATG----- 6096
Qy 619 ArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPhe----- 636
Db 6097 -----GGTGTGGGCAATATTGGGTTCGCTAACACCGCGAGTGGGAATTCGGTATT 6147
Qy 637 -----TyrArgSerThrAspGlyGlyVal 644
Db 6148 GGTGTACCGGTGATATCTACCGGTTCGGTGTTCATATCCGCGAGCGGAATGTG 6207
Qy 645 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal---MetPhe 663
Db 6208 -----GGGTGTGTTAATTCGGGAGCGGTAAATGTGGGTTCCTTT 6246
Qy 664 HisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHis 683
Db 6247 AACTCTGCACCGGGAAC-----TGGGGGGTGTTCATATCGGGGAGTTAT----- 6291
Qy 684 SerThrAsnGlyLysSerTrpSerAlaIleThrGlyValSerSerAlaValAsnVal 703
Db 6292 AACACCGGATCGGTAATTCGGGATTCAGACCGGGTGTTCACGCGGGTGGGTTC 6351
Qy 704 GlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIle 723
Db 6352 AACACGGGTGTGGTCAATCGGGTAGCTACACCGCGAGTTCACCGCGGCGAGGCC 6411
Qy 724 GlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsn 743
Db 6412 AAT-----ACGGCGGTTCACACCGGCGAGTGTCAACACGGGTGTGGTCAACACCGGT 6465
Qy 744 AspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 763
Db 6466 GACATC-----AACACCGGGTGGCCAACTCCGGGAC----- 6498
Qy 764 ArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGly-Al 783
```

```
Db 6499 -----GTCAACACCGCGGCTTCATCTCCGGCACTACAGCAACGGC 6540
Qy 783 aProSerGlySerProSerPro-----SerValSerProSerAlaSerProSerLe 800
Db 6541 GCCTTCTGGCGGGCGACTACACGAGCGCTGCTCGGCTTCTCTACACGACCATCAT 6600
Qy 800 uSerProSerProSerSerSerProSerProSerProSerProSerProSerSerPr 820
Db 6601 CCCGAATTCACTGTCTCGGAACATCCACGCGTCCGCGGCGCGGACCCATCATCTG 6660
Qy 820 oSerSerProSerProSerProSerProSerProSerProSerProSerArgSer---ProSe 839
Db 6661 TCGATCCAATTTCCGGCAATTCCTTGGACCTCAGCGCAACCGGCACATCGGCGGCTTC 6720
Qy 839 rProSer---AlaSerProSerProSerProSerProSerProSerProSerProSerSe 858
Db 6721 ACCATCCCGCGGTGTCATTTCCCGATCAGGTTCCGATCGACCCAGTCTTCGACCTC 6780
Qy 858 rSerProSerProThrProSerSerProValSerGlyGlyValLysValGlnTyrLy 878
Db 6781 GCGCCCATCAGCTCCAGGACATCA-----CGATTCCC 6813
Qy 878 sAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeuGlnValValAsnTh 898
Db 6814 GCGCTGGGACTCGACCCCG-----CAACCGGTGTACCGTGGCGCCGATATTTCAGTCA 8867
Qy 898 rGlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgaspG1 918
Db 6868 GGCTCATCATCGATCCATTAGCC-----TTACGCTGCTGG----- 6904
Qy 918 yGlySerSerThrLeuVal 924
Db 6905 -GGTTCATCAACGTTAATG 6922
```

## RESULT 7

```
US-10-995-561-13368
; Sequence 13368, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13368
; LENGTH: 70513
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13368
```

```
Alignment Scores:
Pred. No.: 0.00283 Length: 70513
Score: 261.50 Matches: 232
Percent Similarity: 31.5% Conservative: 91
Best Local Similarity: 22.6% Mismatches: 333
Query Match: 5.1% Indels: 380
DB: 8 Gaps: 48
```

US-09-917-376-1 (1-957) x US-10-995-561-13368 (1-70513)

```
Qy 86 GlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGly 105
Db 20976 GGCCACACGCTGGAGCGCATCACACGGAGCAGCGCGCA----- 21017
Qy 106 TrpAsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsn 125
Db 21018 -----TCTGCACCCAGCGAGTACTCAC 21038
Qy 126 LysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIle 145
```

|   |   |       |   |                                      |
|---|---|-------|---|--------------------------------------|
| D | B | 21967 | C A G A C T G C A T C C C C G G A A T T G C A G G T G C C C A C T A C T T G C C G A G G C A G C C C T A A T T C A       | 22026                                |
| Q | y | 425   | G l y T r p M e t A s p G l u A l a M e t A l a I l e A s p P r o P h e A s n S e r A s p A r g -                       | Met 441                              |
| D | B | 22027 | C T T T G G G G - G T G C C C G C C A G T C C C C C A C C T T T C A G A C A C A C A G A C A A C T T C C C T G           | 22085                                |
| Q | y | 442   | L e u T y r G l y T h r G l y A l a T h r L e u T y r A l a T h r A s n A s p L e u -                                   | ThrLysTrp 458                        |
| D | B | 22086 | C T G C C C C A C C C T G A C C T T C A C T C C A G T C A A G C A G G A C C A A G G C C C A G G A Y G G G A T G G       | 22145                                |
| Q | y | 459   | A s p S e r G l y G l n I l e H i s I l e A l a P r o M e t V a l l y s G l y L e u G l u T h r A l a V a l             | 478                                  |
| D | B | 22146 | C A G G T G G G G G C A C C A A C A C   | 22166                                |
| Q | y | 479   | A e n A s p L e u I l e S e r P r o P r o S e r G l y A l a P r o L e u I l e S e r A l a L e u A l y A s p L e u G l y | 498                                  |
| D | B | 22167 | - C C C A C A T C T G T G T G C C C T G T G C T G T T T C G G G G G C A G C C G A                                       | 22211                                |
| Q | y | 499   | G l y P h e T h r H i s A l a A s p V a l T h r A l a V a l P r o S e r T h r -   | IlePhe 513                           |
| D | B | 22212 | G G C A C T C A G A G C A T C C A C C T C C C G G A G C C C T C A C A G A Y G G G C A G C C C T G G T G                 | 22271                                |
| Q | y | 514   | T h r S e r P r o V a l P h e T h r T h r G l y T h r S e r V a l A s p T y r A l a G l u L e u A s n P r o S e r I l e | 533                                  |
| D | B | 22272 | C T G G C A C C G G C A C A T T C T G T G G C C T C A C C A G A G T C T C A G T A A G T A G G C C G G G - C T C         | 22330                                |
| Q | y | 534   | I l e V a l A r g A l a G l y -   | SerPheAspProSerSerGln- 545           |
| D | B | 22331 | A T T C T G C G G A G N A G C A A C C C A G G G C A C G T C C G T C T T T C T C C T C C C T K C G T T G G T T C         | 22390                                |
| Q | y | 546   | -   | ProAsnAspArgHisValAlaPhe- 553        |
| D | B | 22391 | T T T C A G C C A G C A G C T G T A T G T T G A G C G C C T A A C A C A C G C T G G C G C T T T C T G G A A T G         | 22450                                |
| Q | y | 554   | S e r T h r A s p G l y -   | LysAsnTrpPheGlnGlySerGluPro 567      |
| D | B | 22451 | G G A A C A C A G G G T G G A C C C A C C C G G C C T C A G A G A C A C T G T T C G G G R A G C A C T -                 | 22509                                |
| Q | y | 568   | G l y -   | GlyValThrThr---GlyGly 574            |
| D | B | 22510 | G G G A A C A C G C A A C C A G T A C T C A G G T G G T G C C G A G G G T G T A C C C T A G G T G G C                   | 22569                                |
| Q | y | 575   | T h r V a l A l a L a s e r A l a a s p G l y S e r A r g P h e V a l T r p -   | 587                                  |
| D | B | 22570 | A G T C A G C A T C G G G C T G S A C A G A G C C A C T C A G A T G T G C A G T C A G G C T G A G T G T G T             | 22629                                |
| Q | y | 598   | - A l a P r o G l y A s p P r o G l y G l n P r o V a l V a l T y r A l a V a l G l y P h e G l y A s n S e r           | 605                                  |
| D | B | 22630 | G A C A G G G T C C T G G A C C C C A -   | GCAGGGAGC 22659                      |
| Q | y | 606   | T r p A l a L a S e r G l n G l y V a l P r o A l a A s n A l a G l n I l e A r g S e r A s p A r g V a l A s n P r o   | 625                                  |
| D | B | 22660 | T G G A G G G C T A T C T A - G T G C C T G C A C C G A G C A G -   | 22693                                |
| Q | y | 626   | L y s T h r P h e T y r A l a L e u S e r A s n G l y T h r -   | 635                                  |
| D | B | 22694 | C C C A C T T T C A T G C G C T T G A T T T G G A C C C G C C T T G C C T T C T G T G G G C C T T A C T G C T           | 22753                                |
| Q | y | 636   | -   | PheTyArgSerThrAsp 641                |
| D | B | 22754 | C T G T C C A C G C T C C A G T T T A C A G G G A G G T G G G A G G C G G T G A G C T C A G A C A G A C A A A           | 22813                                |
| Q | y | 642   | G l y G l y V a l T h r P h e - G l n P r o V a l A l a A l a G l y L e u P r o S e r S e r G l y A l a V a l G l y     | 660                                  |
| D | B | 22814 | G E T T C A C T C C C A T T C A C G G C C A G T G C C G A G A G A G G G G A G C C C T G G G C C C T T C G G             | 22873                                |
| Q | y | 661   | V a l M e t P h e H i s A l a V a l P r o G l y L y s G l u G l y A s p L e u T r p L e u A l a L a s e r S e r G l y   | 680                                  |
| D | B | 22874 | A T G T G C T G G T G G G T C T A G - G G A G G A G T G G C A C A G T C A C T G A G C C G A G G A C T T G G G           | 22932                                |
| Q | y | 681   | L e u T y r H i s S e r T h r A s n G l y S e r S e r T r p -   | SerAlaIleThrGlyVal 697               |
| D | B | 22933 | C T G T T G C A G A G C T G T G G -   | CCCTGGGACCCACCGAGCACTTGTCGTGCC 22986 |

```
QY 698 SerSerAlaValAsnValGlyPheGlyLysSer----- 708
Db 22987 AGCAGAGCTTCAGGTGCTGAGCTCCAGCAGAGCTAACTTGCTCTGGGAAATTAAGTGCCA 23046
QY 709 -----AlaProGlySerSerTyrProAla 716
Db 23047 AGTGGGAGCAGCTGCTGACAGCAGGAGGTGGAACCCAGCAGGCGCAATG----- 23100
QY 717 ValPheValValGlyThrIleGlyValThrGlyAlaTyrArgSerAspCysGly 736
Db 23101 -----GGCTCCAGGAGAGGAGCGCGGGAGCGCGGC 23136
QY 737 ThrThrTrp-----ValleuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp 752
Db 23137 TTGCTTGGCTGCGAGGAGACATTCCTCCCTGACTCACCAGCAGCTGGT----- 23190
QY 753 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg---ArgValTyrIleGlyThrAsn 771
Db 23191 -----GTGACTGAGCTCAGCTCTCCCTGCGAGGTGCAAGCAGGGGTGACT 23235
QY 772 Gly-----ArgGlyIleValTyrGlyAspIle----- 780
Db 23236 GGAATCTTTGATGAACACAGAGCTCTGTTACCGAGGACCTGCTCACCACAGTGTCCC 23295
QY 780 ----- 780
Db 23296 AAAACATTAGCTGCGCATGTTCTCCTGCACTTTTACCGCGCACCCACCTCTCTCCC 23355
QY 781 -----GlyGlyAlaProSerGly----- 786
Db 23356 AGGGAAGGGGCTCCATGCTATCATGACGACCAATTTTCTGCTCTTTTACCTCCACCATCAC 23415
QY 787 ---SerProSerProSerValSerProSerAlaSer-----ProSerLeuSerProSer 803
Db 23416 CTTACTCCCAACCAACCATCACCCCACTACCGCCACCTCCACCATCATCACCCCA 23475
QY 804 ProSerProSerSerProSerProSerPro---SerProSerSerSerProSerSer 822
Db 23476 CCATCACCG---TCATCACCCCCACCATCACCGTATCACCCCCACCATCACCGTATCA 23532
QY 823 SerProSerProSerProSerProSerProSerProSerArgSerProSerProSerAl 842
Db 23533 CCCCACCTCCACCGTATCATCACCCCACTACCA-----CCCCACCATCAC 23583
QY 842 aSerProSerProSerSerProSerProSerProSerProSerProSerProSerPr 862
Db 23584 ATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCC 23643
QY 862 oThrProSerSerPro 868
Db 23644 ATCCCATCCCATCCCT 23662
```

## RESULT 8

```
US-11-052-554A-522/c
; Sequence 522, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 522
; LENGTH: 4146
; TYPE: DNA
```

ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-522

```
Alignment Scores:
Pred. No.: 0.00043 Length: 4146
Score: 260.00 Matches: 220
Percent Similarity: 32.3% Conservative: 99
Best Local Similarity: 22.3% Mismatches: 384
Query Match: 5.1% Indels: 287
DB: 12 Gaps: 43

US-09-917-376-1 (1-957) x US-11-052-554A-522 (1-4146)

QY 20 LeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGly----- 33
Db 3315 CTGGCGCGCTGCGCACCTTCCGCCATTCCTCCGGCTTGGCCACCCCGTGTGCGCGT 3256
QY 34 -----ValleuProIleAlaIleThrAlaSerProAlaHisAlaAlaThr 49
Db 3255 TGCCGAGCGTTCGCCCACTTCCGCCGTTCGCCGCGCGCGCGCGAGATGCTGCCGAC 3196
QY 50 GlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleVal 69
Db 3195 TCCGCTTCTCC-----GCCCTTCCGCCGCGGTGCTGCTGCCGCTGTGCGGTT 3148
QY 70 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 89
Db 3147 GGTGATCCGTCGCGCGCGC----- 3127
QY 90 TrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrp 109
Db 3126 -----GCCACCGCTCACCGCTTGTGGACTTTCAGTATGCG---CGTTGGCGGGCGG 3076
QY 110 GlyTyrAsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAla 129
Db 3075 GGCACCATGCGGTGCGCGGAGTCCGCGCTTACCGCTTACCGCGGTGGGTGCACCGC 3016
QY 130 AlaValGly-----MetTyrThrAsnSerTrpAspProAsnAsp 142
Db 3015 GTCGCGCGCAGTCGCGGTGCGCGCTGTGCTCAGTCGCGCATGATACGATGTGACAGC 2956
QY 143 ---GlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr-ProLeuPr 161
Db 2955 GTCGCGCGCTCACCGCTTACCGCGCTGCGCGCTTCCCGCTGCGCGCGCGATGCC 2896
QY 161 oPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspPr 181
Db 2895 CTGG-----CTGCCA-----CGGTATCCGCC 2875
QY 181 oAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerTh 201
Db 2874 G-----GTAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2831
QY 201 rAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAl 221
Db 2830 TGCTTCGCTGCCAC-----GTTGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2780
QY 221 aAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpVal----- 239
Db 2779 TGGTGGCGGGACGAACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2720
QY 240 -----AlaPheAspLysSerSerSerSerLe 248
Db 2719 GGTTCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2660
QY 248 uGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTr 268
Db 2659 CCGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2600
QY 268 pSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheLePr 288
Db 2599 CGCCGCTGTGTGTTATTCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2548
```

```
Qy 288 oHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGl 308
Db 2547 -----GCCGCGGT-----GCCCGCTTGCCTGCGCGTCG 2519
Qy 308 yGly-----ProTyrAspGlySerSerGlyAspValTrpLysPheSerValTh 324
Db 2518 CGGGTTGCCGCGCTTCCGGGACGGTGTGCTGGTGGCTGCGGCGC---GATGCCG 2462
Qy 324 rSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPh 344
Db 2461 CGGCGCCACCTTCCGCGGTGTCCGCGGACCTCC-----2425
Qy 344 eGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGl 364
Db 2424 -GFTGCTCCGGCGCCGCCACCGATCGGGGACCGGATCCGCGGTGCGCGCATACCGC 2366
Qy 364 nIleSerTrpTrp-----2369
Db 2365 CTTGGCGCGCTTTCGCGCGTCTCCGCGGTCCGCGCGTCCGCGGTGCGCGCCTTCC 2306
Qy 369 oAspThrIleIle-----PheArgSerThrAspGlyGlyAlaThr-TripThrA 385
Db 2305 CGACCGCTGCGCGGGAATCAGCGCTTGCCTCCGCGCACCGCGCGCCACCTGGCGG 2246
Qy 385 rGileTrpAspTrpThrSerTyrProAsnArgSerIleuArgTyrValLeuAspIleSerA 405
Db 2245 CATTCGCGCTTCCGCGCGCGCGCGCG-----2217
Qy 405 laGluProTrpLeuThrPheGlyValGlnProAsnProProVal-----ProS 421
Db 2216 -----GCATTGCGCTTGCCTGCGCGCGCGCGCCACCGTGCCTGCGCGCT 2177
Qy 421 erProLysLeuGlyTrpMet-----AspGluAlaMetA 432
Db 2176 CTCACCTTTGGCGCGCGTCCGCGCGCGCGAGCTCTGTGCCAGTCCCGGTAT-GCGGTG 2118
Qy 432 laIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlat 452
Db 2117 CCGCGCGCGCGCGCAATGCTCCGCGCACCGCGCTACCGCATTACACCGCTCCGCGCG 2058
Qy 452 hrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysG 472
Db 2057 GTTCCGCGCATT-----GGCGCGCGCGCGCGCGCTAGCGCGCGGTGCCACCGCGCG 2007
Qy 472 lyLeuGluGluThr-----AlaValAsnAspLeuIleSerProSerGlyAlaProL 490
Db 2006 CCGCGCTGAAGCGCGTTCGCGCTCCACCCCAAGTTGCGGTGTGCGGTGCCAGCC 1947
Qy 490 euleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProS 510
Db 1946 TTGCGCGCGCGCGGTGCGCGCGCGCGCGCGCTGCGCGCGCGCGCG-----1900
Qy 510 erThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuA 530
Db 1899 -----CCAGTGTGCTGTGCTCCGCGGTGCGCGCGTTCGCGCGCGCGCGCGCG 1851
Qy 530 snProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgH 550
Db 1850 CCGCGCGCGCGCGCGTTCGCGCGCGCGCGCG-----1822
Qy 550 isValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyV 570
Db 1821 -----GCCTTGCACCGAC-----TGGGCGCATTGTGCTGTGTTGCCGCGCA 1779
Qy 570 alThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProG 590
Db 1778 TTGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 1729
Qy 590 lyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAla--S 609
Db 1728 -----GCGGTTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 1686
Qy 609 erGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheT 629
```

## RESULT 9

US-11-205-109-1  
; Sequence 1, Application US/11205109  
; Publication No. US20050287641A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris

APPLICANT: Zazopoulos, Emmanuel  
APPLICANT: Staffa, Alfredo  
TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS  
FILE REFERENCE: 3002-2US  
CURRENT APPLICATION NUMBER: US/11/205,109  
CURRENT FILING DATE: 2005-08-17  
PRIOR APPLICATION NUMBER: US/09/976,059  
PRIOR FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: US 60/239,924  
PRIOR FILING DATE: 2000-10-13  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 88421  
TYPE: DNA  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (2077)..(3078)  
OTHER INFORMATION: ORF 1; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (3118)..(4032)  
OTHER INFORMATION: ORF 2; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (6665)..(5814)  
OTHER INFORMATION: ORF 4; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (7703)..(6693)  
OTHER INFORMATION: ORF 5; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (9691)..(10761)  
OTHER INFORMATION: ORF 7; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (12751)..(10829)  
OTHER INFORMATION: ORF 8; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (13617)..(12802)  
OTHER INFORMATION: ORF 9; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15203)..(13614)  
OTHER INFORMATION: ORF 10; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15591)..(15863)  
OTHER INFORMATION: ORF 11; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (15880)..(19035)  
OTHER INFORMATION: ORF 12; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (19032)..(39713)  
OTHER INFORMATION: ORF 13; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (39713)..(65800)  
OTHER INFORMATION: ORF 14; positive strandedness  
FEATURE:

NAME/KEY: misc feature  
LOCATION: (65826)..(66530)  
OTHER INFORMATION: ORF 15; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (66546)..(67370)  
OTHER INFORMATION: ORF 16; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (67384)..(70059)  
OTHER INFORMATION: ORF 17; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (70099)..(70662)  
OTHER INFORMATION: ORF 18; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (70659)..(71906)  
OTHER INFORMATION: ORF 19; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (73439)..(71964)  
OTHER INFORMATION: ORF 20; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (74216)..(73563)  
OTHER INFORMATION: ORF 21; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (75424)..(74213)  
OTHER INFORMATION: ORF 22; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (75535)..(76464)  
OTHER INFORMATION: ORF 23; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (78110)..(76449)  
OTHER INFORMATION: ORF 24; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (79864)..(78107)  
OTHER INFORMATION: ORF 25; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (81624)..(79861)  
OTHER INFORMATION: ORF 26; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (81909)..(81682)  
OTHER INFORMATION: ORF 27; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (82346)..(82062)  
OTHER INFORMATION: ORF 28; negative strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (82587)..(84446)  
OTHER INFORMATION: ORF 29; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (84481)..(85548)  
OTHER INFORMATION: ORF 30; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (85556)..(86845)  
OTHER INFORMATION: ORF 31; positive strandedness  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (87372)..(86803)  
OTHER INFORMATION: ORF 32; negative strandedness  
FEATURE:  
NAME/KEY: misc feature

51263 AA1GGCGGCACG---GTCGTGGTCGCCCGGGGGGAGCATCGATGCCGCGCTCGGC 51319



```
QY 614 a-----AsnAlaGlnI1 618
Db 52300 CGGTGTGTGCGCGCTTGTCCGAGGGCGGACGCGCGGAGGGCGGTGCGGCGCGGT 52359
QY 618 eArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrAr 638
Db 52360 GCGTCCGCGCGGTG---CGTTGTCTGTC---GCGCAACGCGGTGTGTCTCTGTC 52413
QY 638 gSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly-- 657
Db 52414 CCAGCTGACGCGACGACGACGACCTACACATCCGCGTCCGCTGCACTCGACGCGCC 52473
QY 658 -----AlaValGlyValMetPheHisAlaValProGlyLys----- 669
Db 52474 GCTGATCGGACGCGCTTCCACGCGGCACCTGACGATGTGTGCGCGGACGAGGTGCT 52533
QY 670 -----GluGlyAspLeuTrp----- 674
Db 52534 GCGTACCGTCTTACCGTCCGCGATGGGAGCGGTGGCAACATCTCGACACGCCGTC 52593
QY 675 -----LeuAlaAlaSerSe 679
Db 52594 GGTGAGCGTCCCGTCATCGAGGTGCCCGACGCGGCTTCCGAGGCGGTGCGCGCGC 52653
QY 679 rGlyLeuTyr-HisSerThrAsnGlyGlySer- 690
Db 52654 GCGCGGCACACCTTTCGACCTGAGCGCGGAGATCCCGCTCCGGCGGTGCTCTCGCCAC 52713
QY 691 -----TrpSerAlaIleThrClyValSerSerAlaVala 702
Db 52714 CGGCGCGACCGGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 52769
QY 702 snValGlyPheGlyGlySerAlaProGlySerSerTyrProAlaValPheValGlyT 722
Db 52770 -----GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 52817
QY 722 hrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrTrpValLeuI 742
Db 52818 GGATCCGGGGGAGCGCGCGCC-----TGGACGCGCC 52850
QY 742 leAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGly---AspHis- 760
Db 52851 TGCCCGTCCAGTACGCGCACTACGCCCTGTGTGGCAGCGCGAGTGTGTGTGTGTGTGT 52910
QY 761 -----AlaAsnLeuArgValTyrIleGlyThra 771
Db 52911 ACCCGACAGCGCCATCTCCAGCAGGTGCGCCATTTGGCGCGCA----- 52954
QY 771 snGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerSerProSerProS 791
Db 52955 -----CAGCTCGCGGAGCGCCCGACGAGCTACCGTGTGAGTTACCGGTCC 52991
QY 791 er-ValSerPro-----SerAlaSerProSerLeuSerProSer 803
Db 52992 CCGACACCGCGTCCCGCGGAGCGCACCTACCGCGGCGCACACCGTGTGAGTTACCGGTCC 53051
QY 804 ProSerProSerSerProSerPro---SerProSerProSerSerSerProSerSer 822
Db 53052 CCGCGCGGTGCGCGTTCACCACTCCCGCAACTCGCGCGCGCAACCGCGGTGTACCGTCTTCA 53111
QY 823 SerProSerProSerProSerProSerProSerProSerProSerPro---Ser 839
Db 53112 TGACGTGTCAACCGCGCTTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 53171
QY 840 ProSerAlaSerProSerProSerSerSerProSerProSerProSerProSerSer 859
Db 53172 CCATCGGCGTGTGCGGTGCGCG---GAGCGACCGACCCCAACCTTCGACCACTCTATCGSCT 53228
QY 860 ProSerProThrProSerSer 866
Db 53229 TCTTGTGTCAACACCGCTGTGTC 53249
```

RESULT 10

US-11-052-554A-520/c  
; Sequence 520, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 520  
; LENGTH: 3990  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv

US-11-052-554A-520

Alignment Scores:

Pred. No.: 0.0013 Length: 3990

Score: 250.00 Matches: 222

Percent Similarity: 33.1% Conservative: 93

Best Local Similarity: 23.3% Mismatches: 414

Query Match: 4.9% Indels: 225

DB: 12 Gaps: 42

US-09-917-376-1 (1-957) x US-11-052-554A-520 (1-3990)

```
QY 18 SerLeuLeuAlaAlaThrAlaSerPheAlaValAlaAlaLeuGlyValLeuProIle 37
Db 3327 GCCTTGGCGCGGTGGCGCCTCTGCGCGCTTGGCGCCTCTGCGCGCTACCGGCATC 3268
QY 38 AlaIle-----GlyIleValPheAsnGluGlyAlaProGlyIle 77
Db 3267 GCCTTGGTGTGGCGGAGGAGCGCTTGGCGCGGTGTGGCGCGCGCGCGCGCGCG 3208
QY 45 HisAlaAlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPhe 64
Db 3207 GATGCGACCGTT-----GCCGCGCTTGGCGCGGTGGCGCGCGCGCGCGCG 3172
QY 65 ValAsp-----GlyIleValPheAsnGluGlyAlaProGlyIle 77
Db 3171 GTTACCGCGGTGTACCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 3112
QY 78 LeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrp 97
Db 3111 GCGCGCGTG---ACGCGCGTTCGCGCGCTGCTGTCAG---GATGCTGCC---GGCGCG--- 3064
QY 98 IleProLeuLeu-----AspTrpValGlyTrpAsnAsnTrpGlyTyrAsn 112
Db 3063 ---CCCGCGTTCCTGCGGTGGAGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 3007
QY 113 GlyValValSerIleAlaAlaAspProIleAsnThr-----AsnLys 126
Db 3006 GCGCGTGGCGCGGTGGCGCGCTTACCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCG 2947
QY 127 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly-----Ala 144
Db 2946 GCGGTGGCGCGGTGGCGCGCTTACCGCGGAGACACCGCGGTGGCGCGCGCGCGCGCG 2887
QY 145 IleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeu 164
Db 2886 ACCGCGCGCGCGGTGCTGACTACCGCTTGGCGCGCGGTGGCGCGCGCGCGCGCGCG 2827
QY 165 GlyGly-----AsnMetProGlyArgGlyMetGlyGlyArg 176
Db 2826 GCGCGCGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2767
QY 177 LeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 196
```

[illegible]

|      |    |  |      |
|------|----|--|------|
| 1956 | Db | CGCGTCCCCGCGCTTTGCGCGCGCGCGCCCGCCGCAATTACCAAC-----AAGCCGCGCGTT | 1906 |
| 534  | Qy | --IleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAla    | 552  |
| 1905 | Db | :::     :::     :::     :::     :::     :::     :::     :::    | 1846 |
| 553  | Qy | -----PheSerThrAsp  | 556  |
| 1845 | Db | CGCGGTGCCCGCGTTGCCGCGCGTGTGCGGGTGGCGCGCGCGCGCACCGTCTGCACCGCC   | 1786 |
| 557  | Qy | GlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrVal   | 576  |
| 1785 | Db | GGTGGATCCGCG-----GCCGGCGATCCCAACAGACACCGATGCGCGGTACC           | 1735 |
| 577  | Qy | AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProVal   | 596  |
| 1734 | Db | CGCGTTGCCCGCGTTGCCACCACTCCATCGCGCGCGTATCGAAGTGCCTTGGCACC       | 1675 |
| 597  | Qy | ValTyraValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAla     | 616  |
| 1674 | Db | GTTGCCGCATCACCGCCCATGCGCGCGCGCGCGCGCTTTCGCGCGCGCGCGCGCAC       | 1615 |
| 617  | Qy | GlnIleArgSerAspArgValAsnProLysThrPheTyraAlaLeuSerAsnGlyThrPhe  | 636  |
| 1614 | Db | CATGCTGCCGCTCTGTGGGTGCTCGAAGCGCTTACCGCTTGGCCACCGGCTCC---       | 1558 |
| 637  | Qy | TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSer   | 656  |
| 1557 | Db | ---ACGCCACCGCGGCTCC-----ACGTTGCCGCGCTTGC CGCGCTC               | 1516 |
| 657  | Qy | GlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAla   | 676  |
| 1515 | Db | GGTGCCATCGCGCTGCCCGCCCGAGCGCGTTAAGCGCGTGGCGCGTGGCGCGGTGCC      | 1456 |
| 677  | Qy | AlaSerSerGlyLeuTyHisSerThrAsnGlyGlySerTrpSerAlaIleThrGly       | 696  |
| 1455 | Db | CGCGTTGCCGCGCTT-----ACGCGCGCGCGCGCGACCGCGCGCTGC                | 1408 |
| 697  | Qy | ValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyroAla     | 716  |
| 1407 | Db | TTGGGCTCCGCGCTC---GCCGCGCTTACCGCGAGCGCGCGAGCTCGCGCGCACCGCC     | 1351 |
| 717  | Qy | ValPheValValGlyThrIleGlyGlyValThrGlyAlaTyraArgSerAspCysGly     | 736  |
| 1350 | Db | GTTAGGTGCGCGCCAGAGGCGGGCACCGGGGGCGCGTTCGCGCC-----GGC           | 1300 |
| 737  | Qy | ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyGlyAsnTrpGlyGlnAlaIle    | 756  |
| 1299 | Db | ACCTCGCGCGCGCGCATTTGCCGAC-----CAGCCCGCGCGCGCGCGCGCGCGGT        | 1246 |
| 757  | Qy | ThrGlyAspHisAlaAsnLeuArgValTyriIleGlyThrAsnGlyArgGlyIleVal     | 776  |
| 1245 | Db | ACCGCGGCTTGC-----GCCCATGAAAGTGGCGCGCGTTCGCGCGCGCGCGCC          | 1192 |
| 777  | Qy | TyrGlyAspIleGlyAlaProSerGly-SerPro---SerProSerValSerProSe      | 795  |
| 1191 | Db | GTTGCCGCGCGTGTGGGCTGSCCGCGCGCGCGTGGGACCGATCGTAGGCGG--          | 1134 |
| 795  | Qy | rAlaSerProSerLeuSerProSerProSerProSerSerProSerProSerProSe      | 815  |
| 1133 | Db | -----GTCGCGCGGTGCTCGCGCGCG-----CCGCG                           | 1105 |
| 815  | Qy | rProSerSerProSerSerSerProSerProSerProSerProSerProSerProSe      | 835  |
| 1104 | Db | GCGGGGTACCGCGCTTATCCCGACGACAATCAAGGCACAGAAATTCGCGCGCGCC        | 1045 |
| 835  | Qy | rArgSerProSerProSerAlaSerProSerProSerSerProSerProSerSerSe      | 855  |
| 1044 | Db | GGCCCCGCGGTCCCGCCAAACCCACCGGCGCGCGCGCGCGCGCGCGCGCGCA           | 985  |
| 855  | Qy | rProSerSerPro---SerProThrProSerSerProValSerGlyGlyVally         | 874  |
| 984  | Db | GCGCGCGCGCGCTTCGCGTGCATCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG         | 925  |

QY 874 sValGlnTyrLysAsnAspSerAlaPro 884  
 Db 924 GTGCGGGGACAGTCTCACCAGGCTGCGCCG 894

RESULT 11  
 US-11-121-086-22/c  
 ; Sequence 22, Application US/11121086  
 ; Publication No. US20050266459A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: POULSEN, TIM S.  
 ; APPLICANT: NIELSEN, KIRSTEN V.  
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
 ; FILE REFERENCE: 09138.6000-00000  
 ; CURRENT APPLICATION NUMBER: US/11/121,086  
 ; CURRENT FILING DATE: 2005-05-04  
 ; PRIOR APPLICATION NUMBER: 60/567,570  
 ; PRIOR FILING DATE: 2004-05-04  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SOFTWARE: Patent in version 3.3  
 ; SEQ ID NO 22  
 ; LENGTH: 114801  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-11-121-086-22

Alignment Scores:  
 Pred. No.: 0.0149 Length: 114801  
 Score: 250.00 Matches: 240  
 Percent Similarity: 35.1% Conservative: 103  
 Best Local Similarity: 24.6% Mismatches: 385  
 Query Match: 4.9% Indels: 252  
 DB: 12 Gaps: 48

US-09-917-376-1 (1-957) x US-11-121-086-22 (1-114801)

QY 103 TrpValGlyTyrAsnAsnTyrGlyTyrAsnGlyValValSerIle----- 117  
 Db 60092 TGGGCGGGTGGGGCTTTGGGAATGAAGAGGAGGCGCAGTGTCTGCTGGGCGCAGGCAGG 60033  
 QY 118 -----AlaAlaAappProIleAsnThrAsnLysValTyrAlaValGlyMetTyrThr 135  
 Db 60032 GCAAGTGAGGTGAGCGCTCCAGCCACCAGCAGTGTTCAGCGCCTCCAAATTCCTCCATCC 59973  
 QY 136 -----AsnSerTyrAspProAsnAspGlyValIle-----LeuArg 147  
 Db 59972 TCTCTGGCTCAGCTTTCCTTATGCCGATGGGCGCATCTCTGTGAGGAGGCTTGGA 59913  
 QY 148 SerSerAspGln-----GlyAlaThrTyrGlnIleThrProLeuProPheLysLeuGly 165  
 Db 59912 GGAGCTGATGAAGTCAGGCATGGCTCTGGGCGGTGAGG---GTGGCTGTGAGGGTAGCT 59856  
 QY 166 GlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAappProAsnAsnAspAsn 185  
 Db 59855 GCGTCCCTGCC--- 59844  
 QY 186 IleLeuTyrPheGlyAlaProSerGly-----LysGlyLeu 197  
 Db 59843 ---CTCCCTGCTGGTCCCGCAGCAGTGTCTGTCATCTCTGAGGCGCTCTCAGGCGCAT 59787  
 QY 198 TrpArgSerThrAspSerGlyAlaThrTyrSer---GlnMetThrAsnPheProAsp--- 215  
 Db 59786 GAGAGG-----GGTTCCTCATGGAGTGGGACCTGATTTTTTCCCAAGTGA 59739  
 QY 216 ---ValGlyThr-----TyrIleAlaAsnProThrAsp----- 225  
 Db 59738 TGTCTTGCTCTAAATAACATCTCCAGGATTAATGGTCCAGCCGCCCAACAGCACCAGGT 59679  
 QY 226 -----ThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAsp--- 242  
 Db 59678 GATTGAAAAATGGTTTATGGCCCG-----TGGGTGAAATTCAACTTG 59634  
 QY 243 -----LysSerSerSerSerLeuGlyGlnAlaSerLysThrIle 255

Db 59633 CATCTGAGGTGTATTTCCAAGATGGATTTCTCTCCTGGGAGGCGGCACATTCAC 59574  
 QY 256 PheValGlyValAlaAappPro-----AsnAsnProVal-PheTyrSerArgAspGlyG 273  
 Db 59573 CTAATTAATAATGGCAGCTCTCGACATTCATCACCATAGTGGGGGACACCTGTCTGG 59514  
 QY 273 yAlaThrTyrGlnAlaValProGlyAla-ProThrGlyPheIleProHisLysGlyValP 293  
 Db 59513 GGCACAGGGGCGAGGTGTCTCTGGGTAACTCCAGGTCTCTTAATCTCAGCA----- 59461  
 QY 293 heAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspG 313  
 Db 59460 --GATCTGTGGGGGTGGTCTTAGAGATGCCCTCAGCATGTGAGGCGCCAGGGAGG 59403  
 QY 313 LysSerSerGlyAspValTyrLysPheSerValThrSerGlyThrTyrThr----- 329  
 Db 59402 GGGAGCGGGAGCCCCCAGGAAA-----TGAAACAACTCTAGTGGGTGCCCCAGGGA 59349  
 QY 330 --ArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyL 349  
 Db 59348 GGAGGCTGGAAGGAGG-CCCCAGGCTGAG-----GGTAGCACCCGCG 59308  
 QY 349 euThrIleAspArgGlnHisProAsnThrIleMetValAlaThrThrTyrThrArgIleTyrP 369  
 Db 59307 TAGTGTCTGCCAGGAGCCACAGGCCCCCTGTGAGCAGACTCCAGCAGGCGGCCCTGGC 59248  
 QY 369 roAspThrIleIlePheArgSerThrAspGlyGlyValAlaThrThrTyrThrArgIleTyrP 389  
 Db 59247 CCATCTGTTCGAAGAGGCTGGAGTCTCGGCTGAGCTGG----- 59207  
 QY 389 rpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTyrP 409  
 Db 59206 -----CAGTTGACCTTGGATCTGACTGTCTCCCTGAAA 59173  
 QY 409 euThrPheGlyValGlnProAsnProProValPro---SerProLysLeuGlyTyrMet 427  
 Db 59172 TTCCTTTTACCCTCACTCAGGCTCTCGAGTCCCGCAGCAGCGCCAGCAGGATGGCAG 59113  
 QY 428 AspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu----- 442  
 Db 59112 GAGGAA-----GCCCTTCTCCGACGGGCCCTCACCGACCTCTGCTGAGGGGCCCTTGA 59059  
 QY 443 ---TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTyrAspSerGly 461  
 Db 59058 GCCCAGGCTCTTGTCTCAGCAGCTTGGCCGTGGAGATTGGGAATAITGGAGTTTGGG 58999  
 QY 462 Gly-----GlnIle 464  
 Db 58998 GGATCTCGGTGGCTGAGGCGCTGTGCTCATGAAGGACTCCTTGAGTTTGAAGATGAGCT 58939  
 QY 465 HisIleAla-----ProMetValLysGlyLeuGluGluThrAlaValAsnAsp 480  
 Db 58938 CACCTTCTGCACAGCAATAGGTTCATTGTCTATGGGAACAACACTACCACCAACAAAC--- 58882  
 QY 481 LeuIleSerProProSerGlyAlaPro-----LeuIleSerAlaLeuGlyAspLeuGly 498  
 Db 58881 ---CTCCATCCCGCGGGGCGCAGTGTCTTATCACCATTCTCCTAGTGTAGTACAGG 58825  
 QY 499 GlyPheThrHisAlaAspValThrAlaValProSerThrIle-PheThrSerProValPh 518  
 Db 58824 GGCTCAAAAGGGCGAGGTTCTCAGGTACACAGCAGTGCAGCAACAGAGCAGTGAG 58765  
 QY 518 eThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaG 538  
 Db 58764 GGCTTGGGGGTGTCT-----GCCAGGGGCTG 58738  
 QY 538 ySerPheAspProSerSerGlnProAsnAspArgHisValAla-----PheSe 554  
 Db 58737 AGGTGTGCCCATGCAAGAATCCAGGGGTCCGTGGCGCAGGCGCCCTGCCAAGAGTTGGG 58678  
 QY 554 rThrAspGlyGlyLysAsnTyrPheGlnGlySerGluProGlyGlyValThrThrGlyG 574





Db 207 GCGCGCGTCCCGG- - - - -TTGCTTGTGGCGCGGTCGCGTCCGGCGGCTCC 154  
 Qy 831 rProSerProSerArgSerProSerProSerAlaSerProSerProSerProSerProSe 851  
 Db 153 GCGCGTTCGCGGTCG 94  
 Qy 851 rProSer- - - - -SerSerProSerSerProSerProSerProSerProSerProSe 867  
 Db 93 GCGCGCGGTTGCGTGTGTGTCGCGCGGTTTCGCGCGGTTTCGCGCGGACCGCGCGTCA 34  
 Qy 867 rProVal 869  
 Db 33 GCGGTTG 27  
 RESULT 13  
 US-11-136-527-2622  
 ; Sequence 2622, Application US/111136527  
 ; Publication No. US20050287570A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wyeth  
 ; APPLICANT: Mounts, William M  
 ; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
 ; FILE REFERENCE: 031896-041000 (AM101086)  
 ; CURRENT APPLICATION NUMBER: US/11/136,527  
 ; CURRENT FILING DATE: 2005-05-25  
 ; PRIOR APPLICATION NUMBER: US 60/574,294  
 ; PRIOR FILING DATE: 2005-05-26  
 ; NUMBER OF SEQ ID NOS: 362830  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 2622  
 ; LENGTH: 7231  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-11-136-527-2622

Alignment Scores:  
 Pred. No.: 0.00267 Length: 7231  
 Score: 247.50 Matches: 235  
 Percent Similarity: 32.1% Conservative: 124  
 Best Local Similarity: 21.0% Mismatches: 368  
 Query Match: 4.8% Indels: 391  
 DB: 12 Gaps: 53  
 US-09-917-376-1 (1-957) x US-11-136-527-2622 (1-7231)  
 Qy 86 GlyMetTyrArg- - - - -TyrAspAlaAlaAsnGlyArgTrpIleProLeu- - - 100  
 Db 3121 GGAATATACAGAAACTTCTGCTTCTGGGACAGAGAAAT- - -CAGTGGACTTCTCTTCTGG 3177  
 Qy 101 - - - - -LeuAspTrpValGlyTrpAsnAsnTrpGly- - - - - 110  
 Db 3178 AGGAGAGATCTAGAACTTCTACCTCTGGAATAGATAGGAGCCAGTGCTTCTTCTACTGG 3237  
 Qy 110 - - - - - 110  
 Db 3238 AAGAGGAGTCTAGAGACGCTGCTCTCGAGTAGAGACCTCAGTGGACTTCTTCTTCTGG 3297  
 Qy 111 - - - - -Tyr-AsnGlyValValSerIleAlaAlaAspProLeAs 123  
 Db 3298 AGAAGAAGGTTACAGAAACATCTACCTCTGGAATAGAGGACATAGTGTACTTCCGACTGG 3357  
 Qy 123 nthrAenLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGI 143  
 Db 3358 AGAAAGTCAGAAACTTCTGCTTCTCGAGTGGGAGACTTGAGTGGACTTCCCTCAGGAGG 3417  
 Qy 143 yAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLy 163  
 Db 3418 AGAAAGTCTAGAAACATCTGCTTCTCAGTGTAGAG- - -GATGCACCCAGCTTCTTCTACTGA 3474  
 Qy 163 sLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyuArgLeuAlaValAspProAsnAs 183  
 Db 3475 AAGAGGAGGCTAGAGACTTCTGCCTCTCGAATAGAGGACATCAGTGTGCTTCTTCTACTGG 3534

Qy 183 nAspAsnIleLeu- - - - -TyrPheGlyAlaProSerGI 194  
 Db 3535 AAGAGAAATCTAGAAACTTCTGCTCTGGAGTAGAGGATGTCACTGGACTTCTCTCTGG 3594  
 Qy 194 yLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePr 214  
 Db 3595 AAAAGAAGGCTTAGAGACTTCTGCTCTCTGGAATA- - -GAGGACATAGTGTGTCTCC 3648  
 Qy 214 oAsp- - - - -ValGlyThrTyrIleAlaAsn- - -ProTh 224  
 Db 3649 TACTGAAGCAGAGGCTCTAGAAACTTCTGCTCTCGGGGATATGTCACTGGGATTCCTTC 3708  
 Qy 224 rAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSe 244  
 Db 3709 TGGAGAAGATGGTACAGAAACCTCTACTTCTGGAGTAGAGGCTGTGAGT- - -GG 3759  
 Qy 244 rSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAaAspProAsnAs 264  
 Db 3760 TCTTCTCTTCTGGAGGAGGAGGCTCTAGAAACGCTCTGCTCTGGAGTAGAAGATCTTGGTCT 3819  
 Qy 264 nProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProTh 284  
 Db 3820 TCTTACAGAGATAGTCTAGAGACTTCTGCTCTCAGGAGTAGATGTACTGGTATCTCTTC 3879  
 Qy 284 rGly- - - - -PheIleProHisLysGlyValPheAspProValAsnHi 298  
 Db 3880 TGGAGAGAGAGGACACAGAAACCTCTGTCTCT- - -GGGCTAGGTAGTACCTTAGTGG 3933  
 Qy 298 sValLeuTyr- - - - -IleAlaThrSerAsnThrGly- - - - - 308  
 Db 3934 ACTTCTCTTCTGGACAAGAGCCTAGAGACCTCAGCTTCAGGAGCTGAGGACCTTGGTGG 3993  
 Qy 309 - - - - -GlyProTyrAsp- - - - -GlySerSerGlyAspValTrpLysPhe- - - 321  
 Db 3994 CTGCTCTTCTGGAAAGAGAACTTGTAGGCTCTGCTCTCTGGAGCCCTCGGACTTCTGGCAA 4053  
 Qy 322 - - - - -SerValThrSerGlyThrTrpThrArgIleSerProValProSerTh 337  
 Db 4054 GCTACCTTCTGGAACTCTAGGAAGTGTCAAACTCCAGAGACTAGTGGCTTCTCTCTGG 4113  
 Qy 337 rAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeu- - -ThrIleAspArgGlnHisProA 357  
 Db 4114 ATTTAGTGTGAA- - - - -TATTCTGGAGTGACATTTGGAAGTGGCCCATCTCTC 4161  
 Qy 357 snThrIleMetVal- - - - -AlaThrGlnIleSer- - - - - 366  
 Db 4162 TGGCTCTGCTGACTTTTAGTGACTTCCATCGGCTTCCCAACAGCTCTCTCTTGTGGACAG 4221  
 Qy 367 - - - - -Trp- - - - -TrpProAsp- - - - - 370  
 Db 4222 TACTTAGTGGAGTGATCAGCCACCACCTGCGGAGTGAAGAGGAGGGGACCAT 4281  
 Qy 371 - - - - -ThrIleIlePheArgSerThrAspGly- - - - -GlyAlaThrTrpThrArgIle 387  
 Db 4282 CAGCGTCAGTGGTTCAGGAGAAGAGTCAGGCGCCCTCAGTGTGGAGTAGTGC 4341  
 Qy 387 rpAspTrpThrSerTyrProAsnArgSer- - - - -LeuArgTyrValLeuA 402  
 Db 4342 G-GACATTAGTGTCTCTCTTACGAACTGAACCTCAGTGGCCAAACATCCGATCTCTTG 4400  
 Qy 402 spIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro- - - - - 417  
 Db 4401 ATCTCAGTGGAGAAACATCTGGATTTTGTAGTGTAGGACGCCATTTGGGTCTTCTG 4460  
 Qy 418 - - - - -ProValProSerProL 423  
 Db 4461 GCAGTGTGAGGAGACATCTGGGATTCCTGAGTGTGAGGACGCGCTCAGGAGTCTCTG 4520  
 Qy 423 ysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp- - - - - 439  
 Db 4521 ACACCACTGAGATATCTGAGCTTAGTGGACTGTCTCTCTGGGCAACACAGATGTCTAGTGGAG 4580

[illegible]

```

Db      5421  CAATGTTTCCCGAGGAACTTGGC----- 5443
Qy      749  yrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisalaenLeuArgArgValTyrlleG 769
Db      5444  -----CATGTCTCTTATGACA-----TATATCCCGGCTCTTT---G 5480
Qy      769  lyThrAsnGlyArgGlyIleValTyrglyAspIleGly----- 782
Db      5481  AGCCAGTGGGAAGCCTCAGCATCTGGGACCTTGGTGGACCTGTAAACAATCTTCCCG 5540
Qy      783  -----AlaProSerGlySerProSerProSerProSerValSerProSerA 796
Db      5541  GGTCTGGGTAGAAGCTTCAGTCCCAAGAGGCAGCAGTACCCATCTGTCTTACCTGAGG 5600
Qy      796  laSerProSerLeuSerProSerProSerProSerSerProSerProSerProSerP 816
Db      5601  CTGAGTGGGAGTCTGTGTCGCCCTCAGGCCAGCAGTCAATTGTCTGATTTCCAGATC 5660
Qy      816  roSerSerSerProSerSer-----ProSerProSerProS 829
Db      5661  TGCATGGAATCACTCTGCTCCCGTAAACAGACTGGAGATGACAACCCAGGACCG 5720
Qy      829  erProSerProSerPro-----SerArgSerProSerProSerAlaSerProS 845
Db      5721  AGGTGAGCAGCAACCCATGGACCTTTTCAGGAAGGCACAGGAGGAGATCCGTGTCCAG 5780
Qy      845  erProSerSerProSerProSerSerSerProSerSerProSerProThrProS 865
Db      5781  AAGTGAATGAGAAATCTAGCACTACCTCCGACATAGATCAGGCACATTGAGGTGCCCT 5840
Qy      865  erSerSerProValSerGlyGlyValIysValGlnTyrlsAsnAsnAspSerAlaProG 885
Db      5841  TTGCCAGGCCATGACTTCTGGACAGCAGACTGAATCAGTGCAGAAATGGTCTGAT---- 5896
Qy      885  lyAspAsnGlnIleLysProGlyLeuGlnValValAsnThrGlySerSerValAspL 905
Db      5897  -----CATACCTCAGAGGTGAATGTTTACC 5921
Qy      905  euSerThrValThr-----ValArgTyTrpPhethrargaspGlyGlySerSerThrl 923
Db      5922  TCAGCACCAACCGTCCAGAGTCCAGTGGCCAGTCTACCCAGCACCTTACAGAGACAC 5981
Qy      923  euValTyraenCysAspTrpAlaAlaIleGlyCysGlyAsnIleArgAlaSerPheGlys 943
Db      5982  TT-----CAGAATAATCGGAT 5996
Qy      943  erValAsnProLa-----ThrProThrAlaAspThr 953
Db      5997  CCCCCAATCCCTCATCTATCTCAGGAGAAGAGACCCCAACAGCAGAAACA 6043

RESULT 14
US-11-056-470-2
; Sequence 2, Application US/11056470
; Publication No. US20050261263A1
; GENERAL INFORMATION:
; APPLICANT: SANTI, Daniel V.
; APPLICANT: MYLES, David C.
; APPLICANT: TIAN, Zong-Qiang
; APPLICANT: HUTCHINSON, C. Richard
; APPLICANT: JOHNSON, Robert G., Jr.
; APPLICANT: ZHOU, Yi-Qing
; APPLICANT: FENG, Li
; TITLE OF INVENTION: BENZOQUINONE ANSAMYCINS
; FILE REFERENCE: 300622007510
; CURRENT APPLICATION NUMBER: US/11/056,470
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: US 10/212,962
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/310,779
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/389,255
; PRIOR FILING DATE: 2002-06-14

```



; PRIOR APPLICATION NUMBER: US 60/393,929  
 ; PRIOR FILING DATE: 2002-07-03  
 ; PRIOR APPLICATION NUMBER: US 60/395,275  
 ; PRIOR FILING DATE: 2002-07-12  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 6360  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces geldanus  
 ; US-11-056-470-2

## Alignment Scores:

Pred. No.: 0.00257 Length: 6360  
 Score: 247.00 Matches: 227  
 Percent Similarity: 32.4% Conservative: 77  
 Best Local Similarity: 24.2% Mismatches: 338  
 Query Match: 4.8% Indels: 300  
 DB: 12 Gaps: 49

US-09-917-376-1 (1-957) x US-11-056-470-2 (1-6360)

Qy 24 AlaSerPheAlaValAlaAlaLeuGlyValLeuProIleAlaIleThrAlaSer--- 42  
 Db 3409 GCGTCGTCGCGAGGCGGAGGCTTCGGATCCACCGCGCGTCTGGAGCGGACATG 3468  
 Qy 43 -----ProAlaHisAlaAlaThrThrGlnProTyr 52  
 Db 3469 CACGCGGGCAGCTTCCTCTGCTCCGACCGCGCGGACAGGTGACCTGCTGCGGTC 3528  
 Qy 53 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGlu 72  
 Db 3529 GCCTGGAACACCGTCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3558  
 Qy 73 GlyAlaProGlyIle---LeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 91  
 Db 3559 GCGCGTTCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3603  
 Qy 92 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 111  
 Db 3604 GCCTTCGCTAGCG-----CTGACCGCGGCTCGGCGGCGGCGGCGGCGGCGGCG 3654  
 Qy 112 AsnGlyValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 131  
 Db 3655 GACTCGCTACCTTTCGCGGCGGTCGAGCCCGCGGCGGCGGCGGCGGCGGCGGCGG 3714  
 Qy 132 GlyMetTyrThrAsnSerTrpAspProAsn-----AspGlyAlaIle----- 145  
 Db 3715 GCGCTGTGGACGCTCGCTGGAGCGAGACCTCGCTGCGGCGGCGGCGGCGGCGGCGG 3774  
 Qy 146 ---LeuArgSerSerAspGlnGlyAlaThr-----TrpGlnIleThrProLeuProPhe 162  
 Db 3775 CGCTCGGCGAGTGGCCACCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3834  
 Qy 163 LysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu-----ArgLeuAlaValAsp 180  
 Db 3835 GGCCAGGCGGCGGCTTCGCGA-CGTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3893  
 Qy 181 ProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly----- 196  
 Db 3894 CCTACCG-GACCGC-----CGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3937  
 Qy 197 ---LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAsp 215  
 Db 3938 GCCTGTGGAGGAGATCCAGCGGTGGG-----TCGCGGCGGCGGCGGCGGCGGCGGCGG 3988  
 Qy 216 ValGly-----ThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAsp 232  
 Db 3989 CGCGGCTCGCGGTGTCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4048  
 Qy 233 Ile-----GlnGlyValValTrpValAlaPheAspLysSerSerSerLeu 248  
 Db 4049 ACCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4108

Qy 249 Gly-----GlnAlaSerLys-ThrIlePheValGlyValAlaAs 261  
 Db 4109 GGGTGGCGCTGCTGGATGCCGACGAGCGTCCGAGAACTGCCCGCGGGTGTGCTCCG 4168  
 Qy 261 pProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProG 281  
 Db 4169 GGGACGAGCCCAACT-----GGCGGTGCGGTGCGCGCGGTGTG----- 4208  
 Qy 281 yAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTy 301  
 Db 4209 -----GGTCCACGCTCACCGCGGTGAGCCCGGCTGCGCGTCCGCGCGCAGG 4258  
 Qy 301 rIleAlaThrSerAsnThrGlyGly---ProTyrAspGlySerSerGlyAspValTrpLy 320  
 Db 4259 CGTCTGGCATCTGACATCGCGCGAGTACGGCACCTCGACAATCTGGCGTCTGCCCG 4318  
 Qy 320 s-----PheSerValThrSerGlyThrTrpThrArgIleSerProVa 334  
 Db 4319 ACAGCGCGGACCGCGCACCGCGCGCGGTGCGGA-TCGAGGTCCGCGCGCGC 4377  
 Qy 334 lProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg 354  
 Db 4378 GGGCTCAACT-----TCGCGGATG----- 4396  
 Qy 354 nHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePh 374  
 Db 4397 -----TCCTGGTGGCTCTCGCATGTATCCG 4422  
 Qy 374 eArg-----SerThrAspGlyGlyAlaThrTrp-----Th 384  
 Db 4423 GCGCGTGGGTATCGGCAGCGAGGCGCGGTGTGTGACCGAAATCGTCCGGCGGTC 4482  
 Qy 384 rArgIleTrpAspTrpThrSer-----TyrProAsnArgSerLeuArgTy 399  
 Db 4483 ACGGCTTGGCGGTGGCGACCGGGTGATGGGCTGTCTCTCGGTCTGTGACCGCGC- 4540  
 Qy 399 rValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVa 419  
 Db 4541 -----TGCC-ACCGCGCGCGCGCGCGCGGTATCCGGAT 4574  
 Qy 419 lProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAs 439  
 Db 4575 GCGCGAG-----GGCTGGTGC----- 4590  
 Qy 439 pArgMetLeuTyrGlyThrGlyAla-----ThrLeuTyrAl 451  
 Db 4591 -----TTCGCGACCGCGCGCGGCGGTGCGGTGCGCTATCTGACGCGGCTGTACGC 4640  
 Qy 451 aThrAsnAspLeuThrLysTrpAspSerGlyGlyGln-----IleHisIleAlaProMe 469  
 Db 4641 GTTCGACGACCTCGGGAGGTCCAGCGCGGCGGAGAGGTCTCTGGTGACCGCGCGCGG 4700  
 Qy 469 tValLysGlyLeu-GluGluThrAlaValAlaAsnAspLeuIleSerProSerGlyAlaP 489  
 Db 4701 CGGTGTGGCATGCGCGCGGTCCAGCTCGACAGCACCTTCGGCGCGCACCGTCTCGGCAC 4760  
 Qy 489 rOleuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValP 509  
 Db 4761 CGCCACCCCTCCCA-----AGCACCCACGACCTCCACCGCGGTGGCGGTTC 4805  
 Qy 509 rOserThr-----IlePheThrSerProValPheThrThrGlyThrSerValA 525  
 Db 4806 CGCGAAGCGGTCCCTCCAGCGCGGACCTCGCTTACCGCGGACACCTTCCCGCACCG- 4861  
 Qy 525 sPtyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerG 545  
 Db 4862 -----CCGAGTGTCTCACTCCCTCACCAGCGGACGACATCGACGC 4904  
 Qy 545 lnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGly- 564  
 Db 4905 CTCCTCGGAC-----TTCACACCCCGCGCGCGGTCTCTGGAGATGGGAA 4952

|    |      |   |      |
|----|------|---|------|
| Qy | 565  | SerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAla-----AspGly     | 582  |
| Db | 4953 | GACCGACCTCGGGAGCGCGCGAGTCGGGGGCGGCATCCGGAGTCACTACCGGC           | 5012 |
| Qy | 583  | SerArgPheValTrp--AlaProGlyAspProGlyGlnProValValTyrAlaValGly     | 601  |
| Db | 5013 | GTTCTGATCTCGTGGGAGGCCCGCGGAGCGGGTCTGGCAGCAGTTGGT                | 5072 |
| Qy | 602  | PheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAlaAsnAlaGlnIleArgSerAsp | 621  |
| Db | 5073 | -----GGAGCTGTTTCAGGCGGGCGGATCGAGCC-----                         | 5102 |
| Qy | 622  | ArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAsp    | 641  |
| Db | 5103 | -----GCTGCCGGTACGCGAGTGGGACATACCCCGCGCCC                        | 5138 |
| Qy | 642  | GlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyVal    | 661  |
| Db | 5139 | CGAGGGCTTCGGCTGGATGA-GTCAGGGCGGCATACCG-----                     | 5176 |
| Qy | 662  | MetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeu    | 681  |
| Db | 5177 | -----GCAAGATCGTGTCTCACCC  | 5194 |
| Qy | 682  | TyrHisSer-----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSer     | 699  |
| Db | 5195 | TGCCAGCGCCCTGGACCGCGACCGCTCTCTGTCAACG-----                      | 5236 |
| Qy | 700  | AlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVal    | 719  |
| Db | 5237 | -----GTGCACGGGACCCCTCGCGCCACGATCG-----CCCGCACCTTCTCA            | 5281 |
| Qy | 720  | ValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrp       | 739  |
| Db | 5282 | CCCAGCACG-----  | 5290 |
| Qy | 740  | ValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp    | 759  |
| Db | 5291 | -----GGCACGCCATCTGCTGTGTGTGTCAGCGCC-----GGGGAC                  | 5326 |
| Qy | 760  | HisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp    | 779  |
| Db | 5327 | CGGACGACCTG-----  | 5338 |
| Qy | 780  | IleGlyGlyAlaProSerGlySerProSerProSerValSerProSerAlaSerProSer    | 799  |
| Db | 5339 | -----GCGCCACAGACCTGACCA-----CCGAACCTACCGAAGCTGGCGCCACCGTCC      | 5386 |
| Qy | 800  | LeuSerProSerProSerProSerProSerProSerProSerProSerProSerSer       | 819  |
| Db | 5387 | GCATCACCG-----CCTGGCACACCGCCGACCGGACCACTGCGCGCGCTCTCTCG         | 5437 |
| Qy | 820  | ProSerSerSerProSerProSerProSerProSerProSerProSerProSerProSer    | 839  |
| Db | 5438 | CCGACATCCCGCGGACC-----ACCCCTCTACCGCGGTGGTCCACACGCGCGGGA         | 5488 |
| Qy | 840  | ProSerAlaSer-----ProSerProSerSerSerSerProSerProSerSerSer        | 855  |
| Db | 5489 | CCCTCGACGACGGTGTCTGACCGCGCTACCCCGGACCGCTCGACACCGTCTCTCCGCC      | 5548 |
| Qy | 856  | ProSerSerSerProSerProThrProSerSerSerSerProValSer                | 870  |
| Db | 5549 | CCAAGTTCAGCGCGGTCCATCTCCAGCACTCCACCGCGGACC                      | 5593 |

**RESULT 15**

RESOLUTION 13  
US-11-052-554A-541/c

US 11 052 554A-341/C  
; Sequence 541, Application US/11052554A

; Publication No. US20050288866A1

**GENERAL INFORMATION:**

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

**; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL**

; FILE REFERENCE: 30853/40359A

```

; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 541
; LENGTH: 2352
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
; US-11-052-554A-541

```

Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00197 | Length:       | 2352 |
| Score:                 | 343.00  | Matches:      | 186  |
| Percent Similarity:    | 32.2%   | Conservative: | 55   |
| Best Local Similarity: | 24.8%   | Mismatches:   | 289  |
| Query Match:           | 4.7%    | Indels:       | 222  |
| DB:                    | 12      | Gaps:         | 33   |

US-09-917-376-1 (1-957) x US-11-052-554A-541 (1-2352)

|    |      |  |      |
|----|------|--|------|
| Qy | 251  | AlaSerIysThrIlePheValGlyValAlaAlaAspProAsnAsnProValPheTrpSerArg              | 270  |
| Db | 2175 | GC CGG C C C C A C C C A C C C G C G T G C G A A C C C G C C A A C C C G C C | 2128 |
| Qy | 271  | AspGlyGlyAlaThrTrpGlnAlaValGlyAlaProThrGlyPheIleProHisLys                    | 290  |
| Db | 2127 | ---GGCCCCCGCTGGACAACAGACGCCCGCGTTCACCACGGC---                                | 2086 |
| Qy | 291  | GlyValPheAspProValAsnHisValLeuTyrlleAlaThrSerAsnThrGlyGlyPro                 | 310  |
| Db | 2086 | -----  | 2086 |
| Qy | 311  | TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg                 | 330  |
| Db | 2086 | -----  | 2086 |
| Qy | 331  | IleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrsSerGlyLeuThr                | 350  |
| Db | 2085 | ---CCCGCGCTCCCAACCGAGGAACAAATCC-----GCCGGCGCGCCGCGCGCGCCG                    | 2035 |
| Qy | 351  | IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp                 | 370  |
| Db | 2034 | GCTGCCGAACAGACAGCC-----TCGGCTACCGCGCGCCCGCGTCCCCACC                          | 1987 |
| Qy | 371  | ThrIle---IlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrp                 | 389  |
| Db | 1986 | GGTGTCCAGGTTTTTCGCGTCACCGCGCGCTCCGCGCGCGCGCGCGCGCGGAGCAT                     | 1927 |
| Qy | 390  | ThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAlaGlu-ProTrpIle                  | 409  |
| Db | 1926 | TCCGGCATTCCTCCCGCGCGCCGCCAACCCCGCGCTGCGCATTTAGTGTAGGCGCG---                  | 1872 |
| Qy | 409  | uThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyrTipMetAspG1              | 429  |
| Db | 1871 | -----CCGGTCCCGCGGTGACTCCGCGG-----  | 1848 |
| Qy | 429  | uAlaMetalIleAspProPheAsnSerAspArgMetLeuTyGlyThrGlyAlaThrIle                  | 449  |
| Db | 1847 | ---GCACCGATCAGCCCGCGCAGCAGTCGCGCGGTCCG-CGCGCGCCACCGCGCCCGC                   | 1793 |
| Qy | 449  | uTyAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMe                     | 469  |
| Db | 1792 | CGTCGCACCAACAAGCGCGGTTCACCGCGCCCGCGCGCCACCGCGC---GCCCCACA                    | 1736 |
| Qy | 469  | tValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaPr                 | 489  |
| Db | 1735 | G-----CCGACCGACCCCGCTCCCGCGCGTCCCGCGA  | 1700 |
| Qy | 489  | oLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValPr                 | 509  |



| Result No. | Query |       |        | DB | ID                   | Description        |
|------------|-------|-------|--------|----|----------------------|--------------------|
|            | Score | Match | Length |    |                      |                    |
| 1          | 219.5 | 5.4   | 2736   | 2  | US-09-252-991A-30227 | Sequence 30227, A  |
| 2          | 181.5 | 4.5   | 3892   | 2  | US-09-328-352-5503   | Sequence 5503, A   |
| 3          | 167.5 | 4.2   | 1164   | 2  | US-09-902-540-12627  | Sequence 12627, A  |
| 4          | 166.5 | 4.1   | 1300   | 2  | US-09-902-540-9932   | Sequence 9932, A   |
| 5          | 162.5 | 4.0   | 772    | 1  | US-08-410-784A-5     | Sequence 5, Appl   |
| 6          | 159.5 | 4.0   | 1060   | 2  | US-09-248-796A-14123 | Sequence 14123, A  |
| 7          | 158.5 | 3.9   | 774    | 2  | US-09-346-237-8      | Sequence 8, Appl   |
| 8          | 157.5 | 3.3   | 688    | 2  | US-09-489-039A-9813  | Sequence 9813, A   |
| 9          | 154   | 3.8   | 1548   | 2  | US-09-252-991A-22301 | Sequence 22301, A  |
| 10         | 152.5 | 3.8   | 1751   | 2  | US-09-136-574A-44    | Sequence 44, Appl  |
| 11         | 151.5 | 3.8   | 3472   | 2  | US-09-408-020-4      | Sequence 4, Appl   |
| 12         | 151.5 | 3.8   | 776    | 2  | US-09-346-237-4      | Sequence 4, Appl   |
| 13         | 148.5 | 3.7   | 750    | 6  | 5457037-3            | Sequence 4, Appl   |
| 14         | 148.5 | 3.7   | 751    | 6  | 5457037-5            | Patent No. 5457037 |
| 15         | 148.5 | 3.7   | 776    | 2  | US-09-346-237-7      | Patent No. 5457037 |
| 16         | 147   | 3.6   | 3623   | 2  | US-09-341-461-2      | Sequence 7, Appl   |
| 17         | 146   | 3.6   | 1187   | 2  | US-09-949-016-6513   | Sequence 2, Appl   |
| 18         | 146   | 3.6   | 1749   | 2  | US-09-640-419C-28    | Sequence 28, Appl  |
| 19         | 146   | 3.6   | 2169   | 2  | US-09-949-016-6930   | Sequence 6930, A   |
| 20         | 143   | 3.5   | 894    | 2  | US-08-362-525-22     | Sequence 22, Appl  |
| 21         | 143   | 3.5   | 894    | 2  | US-08-971-692-15     | Sequence 15, Appl  |
| 22         | 143   | 3.5   | 4861   | 2  | US-09-919-497-70     | Sequence 70, Appl  |
| 23         | 142   | 3.5   | 893    | 2  | US-09-605-703B-434   | Sequence 434, Appl |
| 24         | 142   | 3.5   | 1216   | 2  | US-09-134-000C-5130  | Sequence 5130, A   |
| 25         | 141   | 3.5   | 720    | 2  | US-09-296-284-25     | Sequence 25, Appl  |
| 26         | 141   | 3.5   | 754    | 2  | US-09-296-284-4      | Sequence 4, Appl   |
| 27         | 140.5 | 3.5   | 387    | 2  | US-09-253-991A-21962 | Sequence 21962, A  |



```
Query Match      4.2%; Score 167.5; DB 2; Length 1164;
Best Local Similarity 19.8%; Pred. No. 0.0013; Indels 279; Gaps 42;
Matches 173; Conservative 89; Mismatches 333;

QY 6 YTSNVAICGGGFDGIVFNEGAGPILYVRTDGGMYRMDAANGRWIPLLD--DWGWNW 63
Db 179 WAGRNAAGQLG--DGTILDRARPVAVPGLTSTAV---AAGSHVLALRSDGSGVW-AW 231
QY 64 GYN-----GVVSTAAADPINTKNVAAVGMVYNS-----WDPNCGAILRSSD 104
Db 232 GYNALGQLGDDGTTVDRLTPVRVSLGAVVAAAGSYFMSALQSDGTVMTWGGFEGQLGD 291
QY 105 QCATWQITPLPKLGNMPCRGMRGLAVDPNNNDILYFAGSGKGLWRSTDSGATWSM 164
Db 292 GCGVORLSPVRVEGLANITRVAAGSAHALAVRGDGTVMWAGDNGEGQL-----GDGSWADR 347
QY 165 TNFPDVGTYIANPTDTTGVQSDIQGVVAVFADPKSSSLGQASKTI----FVGVADPNPV 220
Db 348 -----FRPVQVPG-----LOGITAVSGGRSHSMALQGGDGTVRAMGYNGYQLGDGT 393
QY 221 FWSRD-----GGATWQAVPGAP 237
Db 394 LTRSRVPALIPGLTGIQALPHSHLHVALHADGTGLRGWYNRFGQLGLGAAGSAVPVQV 453
QY 238 TGFIPHKGVDPVN-----HVLVIATNT-----GGPYDSSGSDVWKFVTSGTWT-RISP- 287
Db 454 RGI-----GRVDRLSVGRATHTLMVRADGTVMWANGENGSGQLGD-----GTSHTRTGPV 501
QY 288 -----VPSTDTANDYFGYS-----GLTIDRQHPNTI-----MVA 316
Db 502 SVPGVPCVRVSAAGTQHSALACDGTVMWANGANARGQLGDGTTTPRVTPLGVEGLRGVVA 561
QY 317 TOISWNPDTIIFRSDDGATWTRIWDTSYNPNRSLRYVLDISAEPWLTFGVQVQNPVPS 376
Db 562 --LAAGDASVALRADGS-----VMSWGG-----NASQLGDGTTLDRLAFA- 601
QY 377 KLGWMEAMADPFNSDRLMYGTGATLYATNDLTKW-----DSGGQIHIAPM-VKG 426
Db 602 HVEHLSNAVPPVPGETHALAVGEDGALWA-----WGANGSQLGDGGGTAPSLTPVRVKG 655
QY 427 LBETAVNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIF---TSPVFTTGTSDVYABL 483
Db 656 LERVA-----SVAAGRAFSAVRDDG-----TAWANGTNPSQLGDGTNHARS 698
QY 484 NPSIIVRAGSDPPSQPNDRHVAFSTGCKNWFQ-----SPPGG-----523
Db 699 VNPQVTSUKGRAMSAHAHVVALSAD-GTVMWGDNTLQGLGDGSSGPTGMVPROVP 757
QY 524 -----VTTGG--TVAASADGSRFVWAPDGPQPVVAVGVFGNSWAASQGVPAQAQIRSDR 576
Db 758 RGVAAVAGEQFTVAVLLDGTARANGSNEYGO-----LGDKTGPQLTPMAVKIEGPK 810
QY 577 VNP--KTFYALNSGTFYRSTGGV-----TFQPVAAGLPSSGAVGVMFHAPVKGEGDLW 628
Db 811 IRPPVTRVAMQHAHVQMSDGTQVTWGDNTFGQLGDGTTTTRAVPLTVYGLASV-----865
QY 629 LAASSGLYHS---TNGSGSWAITGVSSAVNVGFKSAPGSYPA-----VFVV 674
Db 866 VAVSSGAWHSLALLSDGTVMA--WGAN-----GFGQLGDGTSTPRTPVQVAGLDGVVAI 918
QY 675 GTIGVTCAYRSDCGTTWVLINDDQHOYGNWQ 708
Db 919 GSGGTHALAVCSGDSVWTW-----GYNAPGQ 944

RESULT 4
US-09-902-540-9932
; Sequence 9932; Application US/09902540
; Patent No. 683447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
```

```
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 9932
LENGTH: 1300
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-09-902-540-9932

Query Match      4.1%; Score 166.5; DB 2; Length 1300;
Best Local Similarity 21.8%; Pred. No. 0.0018;
Matches 158; Conservative 95; Mismatches 240; Indels 231; Gaps 48;

QY 24 FNEGAPILYVRTDGGMYRMDAANGRWIPLLD---WVGM--NNWYNGVWVSTAAADPIN 77
Db 639 YNPQFGLPATIDLTGAYAGTKVQVRFRVGTDESQGYTCWLLDDLEFNGI-----689
QY 78 TNKVWAAVGMVYNSWDPNCGAILRSSDQCATWQITPLPKLGNMPCRGM--GERL-----131
Db 690 TMTPEFATIA-----PEDGVC-----VNPWPIANAG--PDRSIAPGSELVLYG 729
QY 132 -AVDPNNDNILYFAGSGKGLWRSTDSGATWSQMTNFPDV---GTIANPTDITGGYOSDI 187
Db 730 SAADPEGQVTF-----TWAO--TSGPAVTLAGANTLNFSFTAPQVTES 771
QY 188 QGVWVAVFADPKSSSLGQASKTIIFVGVADPNPVFWSRGGATWQAVGAPTGFIPHKGVF 247
Db 772 TALVFTL--TVSDGVKSTSDSVTVTVALPNP-----PTVNAGLDGIIVERAEY 818
QY 248 DPNVHVLVIATNTGGPYDSSGSDVWKFVTSGTWTRISVPSTDTANDYFGYSGLTIDR 307
Db 819 -----TLSCSASDADGNA-----LTYLWTVQSGTPV--AVKDY-----849
QY 308 QHPNTIMVATQISWNPDTIIFRST--DGGATWT-RIWDTSYNPNRSLRYVLDISAEPWLT 364
Db 850 TPTATFIAPEVT-LDESILVRLTVSDGIATVNDTIVTVTNANRA-----PIVSDTSVA 903
QY 365 FGVPQNPVPPVPSKLGWMD-EAMADPFNSDRLMYG---TGATLYATN-----407
Db 904 FAA-----GTVVTASAVDP-DGDALSYSEWQSGSTVINGADTSALSAPTV 951
QY 408 -----DLTKWDSGGQIHIAPMVK-GLBETA-VNDLISPPSGAPL 444
Db 952 PGSEYFTVTATDGSASAKAVPVTIIDGSLPANSAPTVNAGIDATANAGDWVTLTSGSAS-1010
QY 445 ISALGD-----LGG-----FTHADVTAVPSTIFTSPVFTTGTSDVYABLNPS-----486
Db 1011 -DAEGDITLYHWEQIGGTDVTLTGAD-TLTPS--FTAPSTASGTLTGLFI-LTVSDGTSTT 1065
QY 487 -----LIVRAGSFPDSSQPNDRHVAFSTGCKNWFQSGEPGVTTCGTVAASADGSR--F 539
Db 1066 SDARVIVAA-----DGNANPS--NTAPEADAGESAIV-AEGATVTLNCT-ATDADGDTLVI 1118
QY 540 VNAPDGPQPVVAVGVFGNSWAASQGVPAQAQIRSDRVNPKTFYAL--SNG--TFYRSTDG 596
Db 1119 VMTQ-IGGTPV--TLSDASSLTPFTTAPASS-----DPLTFLMVSDGTATVVDVTSI 1168
QY 597 GVTFOFVAAGLPSGAGVGVMFHAPVKGEGDLWLAASSGLYHSYNGGS-----SWSAITGVS 652
Db 1169 SVTEENVA---PVASARAVL-----SGNOTSATLDGSSDANGDVLTYRTVQVSGPN 1218
QY 653 SAVN 656
Db 1219 ATIS 1222

RESULT 5
US-08-410-784A-5
```

; Sequence 5, Application US/08410784A

; Patent No. 5912413

; GENERAL INFORMATION:

; APPLICANT: MYERS, ALAN M.

; APPLICANT: JAMES, MARTHA G.

; TITLE OF INVENTION: ISOLATION OF SUL, A STARCH DEBRANCHING

; TITLE OF INVENTION: ENZYME, THE PRODUCT OF THE MAIZE GENE

; TITLE OF INVENTION: SUGARY 1

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Weingarten, Schurgin, Gagnebin and Hayes LLP

; STREET: Ten Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 1.5

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/410,784A

; FILING DATE: 24-MAR-1995

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Heine, Ph.D., Holliday C

; REGISTRATION NUMBER: 34,346

; REFERENCE/DOCKET NUMBER: ISU-002XX

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-542-2290

; TELEFAX: 617-451-0313

; TELEX:

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 772 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

; ORIGINAL SOURCE:

; US-08-410-784A-5

Query Match

Best Local Similarity 19.8%; Score 162.5; DB 1; Length 772;

Matches 162; Conservative 90; Mismatches 272; Indels 293; Gaps 42;

QY 51 WIPLLDVGNNWGCYN---GVVS---IAADPINTNKVAAVGMVYNTNSWDNDGAILRSSD 104

DB 108 WGP--NWPYASNWKGSGAGCVSDVDANGDRFNENKLLLDYAEBSQDP----- 155

QY 105 QGATWQITPLPFLKGGNMPGRGMGERLAVDPNNNIIYFGAPSGKGLWRSTDSGATWSQM 164

DB 156 -----LNPSNQ-----GNVFASAHYRTDSDGI----- 178

QY 165 TNPDPDVGTYIANPDTTG-----YQSDIQGVWVAFDKSSSLG----- 203

DB 179 --YAPKGVLPVPSQTGTGTRKTAQKDDVVEVHVRG-----FTEQDTSIPAQVGTYYG 231

QY 204 ---QASKTIFVGVV-----DPNNPV-----FWSRDCGATWQA 232

DB 232 AGLKASYLASLGVTAVERFLPVQETQNDANDVVPNSDANQYWGTYNTENYFSPDRRYAANK 291

QY 233 VPGATG-FIPKGVDPVNHVLYATSNITGGYDGGSGDVWKFVSVTGWTTRISPVST 291

DB 292 AAGGPTAEQAVQAFHAGIKVYMDV-----VYNHTABGGTWTSSDPTTAT 338

QY 292 -----DTANDYF-GVSGLTIDRQHPNTI---MVATQISWMPDTIIFRSTD 332

DB 339 IYSWRGLDNATYYELTSGNYFYDNTGICANENTYNTVAQNILVDSVAYWANTM---GVD 395

QY 333 GGATWTRINDWTSYFNRSRLRYVLDISAPWLTFGVQPPVPSPKLGWMDAEMADPFNS 392

DB 396 GFR-----FDLASVLGNSCLNAVHASA-----PNCPNGGYNFDDAASNVAI----- 436

QY 393 DRMLY-----GTGATLY-----ATNDLTWK-DSGGQIHIAPMVKLEETAVNDL 435

DB 437 NRILREFTVRPAAGCTWICLNGLPSAATR--TSWVDSRRVVRVWVSVPRLRQANL 494

QY 436 ISPPSGAPLISALGDLGTFHADVTAVPSTIFTSVFTTGTSDVYAEINPSIIVR----- 490

DB 495 -----GSMTIYVTQDANDFGSG-----SNLFQSGRSPWNSINFIDVHDGMTLKDVYSC 543

QY 491 --ACSFDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGCTV---AASADGSRF-VWARG 544

DB 544 NGANNSQASYGFSDG-----GTSTNYSWDQGSAG---TGAAVDQRAARTGMAFEMLSAG 596

QY 545 DP-----GQPVVYAVGFGN-----SWAASQG---VPANAQIIRSDR---VNEK 580

DB 597 TPLMQGDEYLRITQCNNAYNLDSSANWLTVSWTTDQSNFTYFAQLIRLSARHIPRES 556

QY 581 TFYALNSGTFTYRSTDDGGVTFQFVAAAGLPSGAGVGMFHAVPGKEGDLMLAASSGLYHSTN 640

DB 657 SWYSGSLTWY-----QP-----SGAV-----ADSNYWNNTS 683

QY 641 GGSSWSAITGVSSAVNVGFGKSAPGSSYPAPVWV--GTIGGVTGAYRSDDCGTTWVLND 698

DB 684 NYAIAYAINGPSL-----GDSNDSIYVAYNGWSSSVTFTLPAPPSGTQWYRVTD 732

QY 699 DQHQVGNWQAITGDHANLRVYICTNGRGIVYGDIG 735

DB 733 T-----CDWNDCASTFVAPGSETLIG--GAGTTYGCGG 763

#### RESULT 6

US-09-248-796A-14123

; Sequence 14123, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 14123

; LENGTH: 1060

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-14123

Query Match

Best Local Similarity 4.0%; Score 159.5; DB 2; Length 1060;

Matches 147; Conservative 68; Mismatches 267; Indels 223; Gaps 33;

QY 7 TWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGG-----MYRWDAAANGRW--IPL-LDW 58

DB 142 TYKNVPAGYRPFVDAYI-----SATDWSVTLSVANEYTCAGYVQWQAPPTLRWT 191

QY 59 GWNW--WGYNGVSIADAPINTNKVAAVGMVYNTNSWDNDGAILRSSDQATWQI--TPLP 115

DB 192 GYRNSAGSNGIVVAITRTVTDSTAVT---TLFPDEN-----RDKTKTIELKPTIP 241

QY 116 FK-----LGGNMPGRGMGERLAVDPNNNIIYFGAPSGKGLWRST--DSGATW 161



Db 242 TTTTTSYGVVTSYLTAKTAP---IGETATVIV---DIPYHTTTTTSKWTGTTTSTTH 295  
Qy 162 SQTNPFP---VGYIANPT-DTGYQSDIOGVVWVAFDKSSSSLGQASKITFVGVADPN 217  
Db 236 TNPSTDIDTVIQVLPNPVTVTTEYSQ-----SPATTTITGPPGNTDVLIREPP 348  
Qy 218 NPV-----FWSRD---GGATQAVCAPGTFPHKGVDPVNVHL-----YIATNT 261  
Db 349 NHTVTTEYSVSTTSTFAPPGTDSVI-----IKEPPNPVTVTTEYSVSTTITTV 404  
Qy 262 GPPYDG-----SSGDVWKFVSSTGTWTRISPVSTD-----TAN 295  
Db 405 TAPPGTDTVLIREPPNHTVTTEYSQSYTT-TTIVAPPGTDSVLIIREPPNPVTVT 463  
Qy 236 DYGVSGLT-----IDRQHPNTIMVATQISW-----PDT 325  
Db 464 EYWSQYATTTTITAPPGTDTVLIREPPNHTVTTE--YWSQSYATTTTITAPPGTDT 521  
Qy 326 IIFSTGGATWTRINDWTSYPNRSLRVLDISAEPLTFGV-----QNPVPSPKL--- 378  
Db 522 VLIREPPNHTVT---TEYSQSYTTTITVIAAPPGTDSVLIKEPPNPVTVTTEYSQ 576  
Qy 379 -----GWMDEAMATDPNSDRMLYGTGATLYATNDLTKWDSGGQIHIA----- 421  
Db 577 SYATTTITAPPGTDTVLIREPPNHTVTTEYSQSYATTTTITAPAGGTDVLIIREPP 636  
Qy 422 -PMVKGLE-----ETAVNDLISPPSGAPLISALGDLGGFTHADVTAPSTIFTSVPFTG 475  
Db 637 NPTVTTEYSQSYTTTNTVTAPOG-----TDSVLIIREPPNPVT 676  
Qy 476 TSDVYAEINPSIIVRAGSDFPSQPNDRHVAFTDGGKWFQSGPGVT----- 525  
Db 677 TTEYSQSYATTI-----TVTATPGGTDVLIIREPPNPVTVTTEYSQS 721  
Qy 526 --TGTVAASADGRFVWAPGDPQPVVYAVFGNSWAASQVPA 568  
Db 722 YATTTTAPPGTDTVLIREPPNPVTVTTEYSQSYATTTTVA 766  
RESULT 7  
US-09-346-237-8  
; Sequence 8, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:  
; APPLICANT: Bisgaard-Frantzen, Henrik  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629.200-US  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER FILING DATE: 1999-07-01  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 8  
; LENGTH: 774  
; TYPE: PRT  
; ORGANISM: Favobacterium odoratum  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(774)  
; OTHER INFORMATION: Isoamylase  
US-09-346-237-8  
Query Match 3.9%; Score 158.5; DB 2; Length 774;  
Best Local Similarity 18.8%; Pred. No. 0.0035;  
Matches 158; Conservative 103; Mismatches 265; Indels 313; Gaps 44;  
Qy 76 INTNKVWAA-----VGMVTNSWDPNDGAILRSSDQCATWQITPL 114  
Db 35 INPNKLGAVDATKANVTFKYSSKATRIELYLSTATGSAEKAKYVMTNSGGIWSVT-I 93

Qy 115 PFKLGNNPGRGMRGLAVDPNDNDNLLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYI 174  
Db 94 P---TSTLSQGLG-----GTLYYGYRAWGNW---PYNASWTKGSSL----- 130  
Qy 175 ANPTDITGYQSDIQG-----VWVAFDKSSSSLGQASKITFVGVADPNP 219  
Db 131 -----GFISDVDAAGNRFNPNKLLSDPYALELSHDPTTATM-----TNGS 170  
Qy 220 VFWSDGGATQAV---PGAPTGF-----PHKGVDPV---NHLVIATSNIG 262  
Db 171 IYAS---GATYRNDSSGAPKIVLAGDTQATGTFKTRALKDDVVEAHVRGLTWNDS 227  
Qy 263 -----GPYDSSGDVWKFVSSTGTWTRISPV-----PSTDANDYFGYSLTI 305  
Db 228 ITAAYRGYTKGAGLKAALAAALGVAIEFLPVQETQNDTNDNDPSSTSGDNYWGYMTLNY 287  
Qy 306 ---DRQ-----HPNTIMVATQISWPDITIIFRSDGGATWTRIW 341  
Db 288 FAPDRRYAYDKTPGGTREFKEMVKAFHDNGIKVLVD-----VYNHTGEGGAMSPT- 339  
Qy 342 DMTSYPNRSLR-----YVLDISAE-PWLTGCVQNPVPVS-----PKLGMDEAM 385  
Db 340 DKTTNITSFRGLDNPYYLSLTADFQNSWDNTGVGNGYNTRNTIAQNLIVDSLAWRDKL 399  
Qy 386 AIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKLEETAVNDLIS--PP--- 439  
Db 400 GVDGYRFDLASVLGNSCQHCFCNFKWDAG-----NALNRIVAEPLPRPAT 445  
Qy 440 ---SGAPLIS---ALG-----DLGGF-----THADVTAPSTIFTSVPFTTGTSDVYA 481  
Db 446 GSGVDLLIAEPWAIGNSYQVGGFPGSWAEWNGAYRDVVRQAQNLKLSVAITTG----- 499  
Qy 482 ELNPSIIVR-ACSFPPSSQPNDRHVAFTDGGKWFQG---SEPGVTTTGTVAASADGS 537  
Db 500 ---QMATRFAGSSD-----LYDGRKPHSVNFIATHDGFTLKDLYSCNSKN 545  
Qy 538 RFVW---APGDPQPVVYAVFGNSW---AASQ-----GVP-----ANAQ 571  
Db 546 NOVWPGFSDGGE-----DNNNSWDQGGIADQKAAANGMALMLLSAGVPMIVGDEA 599  
Qy 572 IRS-----DRVNPRTFY-----ALNSGTFYRSTD-G 596  
Db 600 LRSMCNPNPNLDSSANLWNSRTTDQNNFQSFKAIAFRKAHPALRPANFYSSVDNN 659  
Qy 597 GVTFOFVAAGLPSGAVGVMP-----HAPQK-EGDLMLAASSGLYHSTNGSSWSAIT 649  
Db 660 GNVMEQLRWFKPDGGVADATYFENDANNHAIARWIDGSEFQDTASAIYVAHN---AWSA-- 714  
Qy 650 GVSSAVNVGFGKSAPGSSYPVAVVVGTCIGVTGAYSDDCGTTWVLINDDOHOYGNWQ 708  
Db 715 -----QVNFTLPWFGAGKSWYRVTDTCGMWASGASQVQAPGSE-ALVGGENTAYGLGCR 766

## RESULT 8

US-09-489-039A-9813  
; Sequence 9813, Application US/09489039A  
; Patent No. 6610836

## GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9813  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9813

Query Match 3.8%; Score 157.5; DB 2; Length 688;  
Best Local Similarity 22.6%; Pred. No. 0.0035;  
Matches 111; Conservative 53; Mismatches 184; Indels 143; Gaps 24;

QY 60 WNWG--YNGVSTADPINTNKVAAVGMVTSWDPNDGAILRSSDQATQITPLPFK 117  
DB 171 WAHGNTHGDRFAALDQINKN-----VNLQVAVAHATGDIPOSNGAEDQNTPLQ-- 224  
QY 118 LGGNMPGRGMBRLAVPNDNINLYFCAGSKGLWR--STDGATW-----SOM 164  
DB 225 -----IGDTLYVCTPYSKVLALDSDGKEMRYDSKSSPNWQRCRGLGYADSQA 275  
QY 165 TNFPDVT-----YIANPTDT-----TGVSQDI-----QGVVWVAFD 196  
DB 276 QTAPASGTQPAACSRRLFLPTIDARLIAIDATGKLCENFGDGGIVDLVSGMGEVKAGY 335  
QY 197 KSSSLGQASKTIFVG--VAD--PNNPVFWSRDGGATWQAVCAPGTFPHKG-----VF 247  
DB 336 QQTSTPLVAGNVVVGGRVADNVTGEP-----PGVVRADFVHTGKLAWAW 381  
QY 248 DPNVHLYIATSNTPGPDGSSGDVWKFVSFTSGT---WTRIS-----PVSTDTAND 296  
DB 382 DPGNAL-----TGVPPEQ-----TYTRGTPNVWSAMSVDKLNLIYLPNGNATPD 428  
QY 297 YFGVSGLTIDRQHPNTIMV--AT--QISWMPDTIIFRSTDCGATWTRWDWTSYPNRSR 352  
DB 429 FFGERTALDDKYSSSIVAVDATTTQGVWHFQT-----THDLWDF-DLPSQPLL 477  
QY 353 Y-----VLDISAEPWLTFCGVPNPVSPKLGWDEMAIDPFN-----SDRM 395  
DB 478 YDLPDGGKGTTPVLVQTSKQGMIFMLNRETGEPAKV-----EERPVPAGNVKERYPTQ 533  
QY 396 LYTGATLYATNDLTWDSGQIHIAPMVKGLE--ETAVNDLISPPSGAPLISALDGLG 453  
DB 534 PYSVGMPMIGNQILTSDMWGATPIDLLLCRIQFKEMRHOGVFTPPGEDRSLOPPGLGG 593  
QY 454 FTHADVTAPVS 464  
DB 594 MNWGSVSLDEN 604

## RESULT 9

US-09-252-991A-22301  
; Sequence 22301, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22301  
; LENGTH: 1548  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22301

Query Match 3.8%; Score 154; DB 2; Length 1548;  
Best Local Similarity 20.7%; Pred. No. 0.019;  
Matches 188; Conservative 87; Mismatches 270; Indels 364; Gaps 53;

QY 73 ADPINTNKVAAVGMV---TNSWD-----PNDGAILR----- 101  
DB 319 SDVROGRNWALAEMLPAGSQWSMRLVAGADLAADNRLVRPDSSASLNLDATHYQAKI 378  
QY 102 -----SSDQATWQITPLPFKLGNNPGRGMRGLAVDPNNDNLYFGAGSGKLWR 153

DB 379 EQSSGGLVFTDQATDWTGP-----GTFVDESNEWI-----CGLGPHYC 416  
QY 154 STDGATWSOMTNPFPDVTGTYIANPTDTTGYQSDIQGVVWVAFDKS--SSSLGQASKTIFV 211  
DB 417 AEPRTWAP-----GNLYGMPAGTAIGBDLW--WCSVDPLSLCIENLCKT----- 460  
QY 212 GVADPNPNVF-----MSRDG--GATWQAVPGAPGTFPHKGVPDP 249  
DB 461 -VVTQNLFSVLRTGTGDLDLASAGNLQWSPYGVYTAGTQAADVA-TGFNQPRGLEF- 517  
QY 250 VNHVL-----YIATS-----NTGGPYD--GSS--GDVWKFVSFTSGTWTTRISVP 289  
DB 518 -GSVLGAGGADYEVLSTSQYQAYPEHGGNLDIAVGGDVGVGQWAEKLTSS--DIRPLP 574  
QY 290 STDANDYFYGSLTIDRQHPNTIMVATQISWMPDTIIF-RSTDCGATWTRWDWT----- 344  
DB 575 PSAAVGNWLRQG-SADREGVPT-----AWWNFGSVVRAEGDAPY--LVGTGTGFT 624  
QY 345 -----SYPN-----BSLYVLDISAEPWLT----- 364  
DB 625 LGGNLSMRTGGDAGNIAPRGDGSIPSSGNLNPSSQGLVAVAGTGRLTSDCALQLGGG 684  
QY 365 -----FGVQPNPVPSPKLGWDEMAIDPFNSDRM--LYGTG-----AT 402  
DB 685 DLNVRIGGEVNP-----SREARATQYSSSGFDGLYSGGTHDLOGALINLRGSAS 735  
QY 403 LYATNDLTWDSGQIHIAP--MVKGLETAVNDLISP-----PSGAPLIS----- 446  
DB 736 LY-----SGALGIDPRYDILLRDPAEVRSRDAFSPTLASSTGGTILVAGDTGMRL 786  
QY 447 -ALGD--LGGFTHADVTAVPSTI--FTSP-----VFTGTSTVDYAEINPSII 488  
DB 787 ETRGDLVLGVTDPGRVGVNPTVGTAPDGSVYQGGIGWFLWTAHTSID-----LF 839  
QY 489 VRAGSFDPSSQ-----PNDRHVAFTSD-----GKKNWFQSS 520  
DB 840 AAGGNLTPTQLVETATNAIPMAGENLSPSDGRFTYPSIVRAAAPRGSIYLGPPSGVMGV 899  
QY 521 PGVYTT--GGTVAASADGSRFVWAPGDPGVVYVAVGFGNSWAASQGVANAQI----- 572  
DB 900 SLNVSTTYSLLAPLSINLELELLA-GDS-----IYAGGY--SVQSSGADPANLPISWTPA 952  
QY 573 ---RSDR--VNPKTFYALSNGTFYRSTGGVTFQPVAAAGLPSSGAVGMFHAVPGKEGDL 627  
DB 953 FAGYSDAALNP-----IAGNG-----SPGN--PAVIGGLP-----LFYFGPDSAASL 994  
QY 628 WLAASSGLYHSTNGSSWSAITGVSSAVNVGVFGKSAPGSSYPVFFVGTIGVTVGAYRSD 687  
DB 995 ARDLQPARFYALTGD-----IVGLNSGAQIRFGQA-----GNRAGQWYE 1035  
QY 688 DCGTTWVLINDDQHOYGN-WQOAI-----TGD-----HANLRVYIGTNGRGIVY 731  
DB 1036 GAGPVNMRAGRDIVASGTPLGQRISAFPSQISTDASFTGNLFVHDNDLNLVQAGRDILY 1095  
QY 732 GDIGGAPSG 740  
DB 1096 GNFVAGPG 1104

## RESULT 10

US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for

360 EPWLTFGVQNPVPSPKLGW-----MDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWD SG 415

```

Query Match          3.8%; Score 152; DB 2; Length 3472;
Best Local Similarity 21.0%; Pred. No. 0.083;
Matches 172; Conservative 100; Mismatches 244; Indels 304; Gaps 50;

Qy 96 DGAILRSSDQAGTWQITP-----LPFKLGGNMPGMRGLAVDPNDNNILYFGAPSKGL 151
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 986 DGTWFLSD-GPTGRVVPYSLNIPFDISSAAPG-----GFVIVPGVSDIAFSA-DGRNM 1038

Qy 152 WRSTDGATWSQMTNFP-DVG-----TYANP--TDTTGYQSDIQGVVVAF 195
||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 1039 LVADETGGIHRYLARSFVPIGTDFIKSLGFVEFTPSAAPRVODLAGIAFSDHGMILAA 1098

```



5457037-5  
 ; APPLICATION NUMBER: US/08/1,797  
 ; FILING DATE: 08-JAN-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 749,621  
 ; FILING DATE: 19-AUG-1991  
 ; APPLICATION NUMBER: 224,114  
 ; FILING DATE: 25-JUL-1988  
 ; SEQ ID NO:3  
 ; LENGTH: 750  
 5457037-3  
  
 Query Match 3.7%; Score 148.5; DB 6; Length 750;  
 Best Local Similarity 20.7%; Pred. No. 0.018;  
 Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;  
  
 QY 27 GAPGILYVRTDGGMYRWDANGRWIPLLDWVGNNWNGYGVVIAADPINTNKVAAVG 86  
 DB 70 GITGAVY-----YGYRAGNPNWPYASNMGKSGQAGFVSDVDANGDRFNPKNLLDPY 121  
  
 QY 87 MYTNSWDPNDAILRSSDQATWQITPLPKLGGNMPGRGMRGLAVDPNNDNLLYFGAP 146  
 DB 122 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 156  
  
 QY 147 SGKGLWRSTDGATWSQMTNFPDV--GTYYIANPTD--TTGYQSDIQGVVWVAFDKSS--SS 201  
 DB 157 KGVVLVPSTQSTGKTPRAQKDDVIEVHVRGFTQDTSIPAQYRGTYGAGLKASYLAS 216  
  
 QY 202 LQOASKTIFGVAD-----PN---NPVFW-----SRDGGATQAVPGAPT-G-P 240  
 DB 217 LG-VTAVEFLPVQETQNDANDVVPNSDANQNYGMYTENYFSPDRRYAYNKAAGGPTAEF 275  
  
 QY 241 IPHKGVFDPVNHVLYIATSNLTGGPYDGSDDVWKFVSTGCTWTRISPVPSST----- 291  
 DB 276 QAMVQAFHNAGIKVYMDV-----VYNHTAEGGTWSSDPTTIATYSWRGLDN 322  
  
 QY 292 -----DTANDYF-GYSGLTIDRQHPNTI---MVATQISWMPDTIIFRSTDGATWTRIW 341  
 DB 323 TTYEYLTSGNOYFYDNTGIGANFYNTVAQNLIVDSLAYWANTM---GVDGFR-----P 374  
  
 QY 342 DWTS-YPNRSLRYVLDISAEPMLTFTGVQPNPPVPSPKLGWDEMAID---PNSDRML 396  
 DB 375 DLASVLGNSCLNGAYTASA-----PNCNGYFNDAADSNVAINRILREFTVRPAA 425  
  
 QY 397 YGTGATLYATNDLTWKDSSGQIHI-APMVKGLBE-----TAVNDLISPPSGAPLI 445  
 DB 426 GSGGLDLFA---EPWAGNSYQLGGFPQGWSEWNGLPDLSLRQAQNEL-----GSMTI 476  
  
 QY 446 SALGDLGGTHADVTAVPSTIFTSPTFTGTSVDYAEINPSIIVR-----AGS 493  
 DB 477 YVTQDANDFSGS-----SNLFQSSGRSPWNSINFDVHDGMTLKDQVYSCNGANNQAWP 530  
  
 QY 494 FDPSPQNDHRHVAFTDGGK---NMFQSGEPGVTTGGTV---AASADGSRF-VWAPGD 545  
 DB 531 YGFS-----DGGTSTNYSWDQMSAG---TGAANORRAARTGMAFEMLSAGT 575  
  
 QY 546 P-----QPPVYVAVGFGN-----SWAASQGVPAQAIRSDR-----V 577  
 DB 576 PLMOGGDEYLRITLQCNNAYNLDSSANWLTYSWTDDQ---SNFTYFAQRLIAFRKAHPAL 632  
  
 QY 578 NPKTFVALSNGTFYRSTDGVTQPVAAGLPSSGAVG-----VMFHAVPGKEGD 626  
 DB 633 RPSWYSGSGLTWY-----Q-----SCAVADSNYWNNTSNYAIAYANGPS-- 674  
  
 QY 627 LWLAASSGLYHSTNG-----GSSWSAIT-----GVSSAVNVGFGKSAPG 665  
 DB 675 --LGDNSIYVAYNGSSSVFTPLAPPSTQYRVTDTCDMNDGASTFV-----APG 725  
  
 QY 666 SSYPVAVVGTIGGVTGAYRSDCGTTWVLINDDOHQYGNWGOAI 710  
 DB 726 SE-----TLIGG-----AGIT-----YGCCGQSL 744

Db 676 --LGDSNSIYVAYNGWSSSVFTLPAPSGTQWYRVTDTCNDNGASTFV-----APG 726  
QY 666 SSYPVAVFVGTIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQAI 710  
Db 727 SE-----TLIGG-----AGTT-----YGQCGQSL 745

## RESULT 15

US-09-346-237-7  
; Sequence 7, Application US/09346237A  
; Patent No. 6265197  
; GENERAL INFORMATION:  
; APPLICANT: Bisgard-Frantzen, Henrik  
; TITLE OF INVENTION: Starch Debranching Enzymes  
; FILE REFERENCE: 5629.200-US  
; CURRENT APPLICATION NUMBER: US/09/346,237A  
; CURRENT FILING DATE: 1999-07-01  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: PA 1998 00868  
; EARLIER FILING DATE: 1998-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 776  
; TYPE: PRT  
; ORGANISM: Pseudomonas species SMP1  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(776)  
; OTHER INFORMATION: Isoamylase  
US-09-346-237-7

Query Match 3.7%; Score 148.5; DB 2; Length 776;  
Best Local Similarity 20.7%; Pred. No. 0.019;  
Matches 171; Conservative 88; Mismatches 275; Indels 291; Gaps 47;  
QY 27 GAPGILVVRTDIGMYRWDANGRWIPLLDVGVNNGYNGWSIAADPINTNKVWAAVG 86  
Db 96 GITGAVY-----YGRANGPNWPYASWNGKSGQAGFVSDVDANGDRFNPKNLLDPY 147  
QY 87 MYTNSWDPNCGAILRSSDQAGATQITLPLFKGGNMPGRCMGERLAVDPNNNDILYFCAP 146  
Db 148 AQEVSQDP-----LNPSNQ-----NGNVFASGASYR-----TTDSGIY--AP 182  
QY 147 SGKGLWSTDSGATWSQNTNPPDV--GTYYIANPTD--TTGYOSDIQGVVWVAFDKSS--SS 201  
Db 183 KGVVLVPESTQSTGTPKTRAQKDDVIYEVHVRGFTQDTSIPAQYRGTYGAGLKASYLAS 242  
QY 202 LGQASKTIFVGVAD-----PN---NPVFW-----SRDGGATWQAVPGAPTQ-F 240  
Db 243 LG-VTAVEFLPVQGTQNDANDVNSDANQYWGMYMTENYFSDRRYAINKAAGPTAEF 301  
QY 241 IPHKGVDPVNHVLYIATSNTPGPDGSGDVKFVSFTSGTWTRISVPVST-----291  
Db 302 QAMVQAFHNAGIKYMDV-----VYNHTAEGGTWTSDDPTATIYSWRGLDN 348  
QY 292 -----DTANDYF-GYSGLTIDROHPNTI---MVATQISWWPDTIIFRSTDCGATWTRIW 341  
Db 349 TTYVELTSGNQYFYDNTGIGANFNNTYNTVAQNLIIVDSLAYWANTW---GVDGFR-----F 400  
QY 342 DWTS-YPNRSIRYVLDISAEPWLTFGVQPNPPVPSPKLWMDAEMAD-----PNSDRLM 396  
Db 401 DLASVLGNSCLNGAYTASA-----PNCPNGYNPDAADSNVAINRILREFTVRPAA 451  
QY 397 YGTGATLYATNDLTKWDSGGQIH-IAPWVKGLEE-----TAVNDLISPPSGAPLI 445  
Db 452 GSGSLDLFA-----EPWALGNSYQLGGFPQGSWNGLFRLDSLRQAQNEL-----GSWTI 502  
QY 446 SALGDLGGFTHADVAVPSTFTFTGTSVDYAEINPSIIVR-----AGS 493  
Db 503 YVTQDANDFGS-----SNLFQSSGRSPWNSINFIDVHDGMTLKDVIYSCNGANNSQAWP 556

QY 494 FDPSSQPNDRHVAFTSDGK-----NWFGSEBEGVTTGGTV---AASADGSRF-VWAPGD 545  
Db 557 YGPS-----DGGTSTNYSWDQMSAG---TGAADVQRRRAARTGMAPEMLSAGT 601  
QY 546 P-----GQPVVYAVGFGN-----SWAASQGVPAQAIRSDR-----V 577  
Db 602 PLMQGGDEYLRTLQCNNNAYNLDSANMLITYSWTTDQ---SNFYTFAQRLIAFRKAHPAL 658  
QY 578 NPKTFYALSNGTFFYRSTDCGVTFFQVVAAGLPSSGAVG-----VMFHAVPGKEGD 626  
Db 659 RPSSWYSGSLTWY-----QP-----SGVADSNYWNNTSNYAIATVAINGPS-- 700  
QY 627 LWLAASSGLYHSTNG-----GSWSAIT-----GVSSAVNTVGFSGKSAPG 665  
Db 701 --LGDSNSIYVAYNGWSSSVFTLPAPSGTQWYRVTDTCNDNGASTFV-----APG 751  
QY 666 SSYPVAVFVGTIGVGTGAYRSDDCGTTWVLINDDOHQYGNWQAI 710  
Db 752 SE-----TLIGG-----AGTT-----YGQCGQSL 770

Search completed: March 2, 2006, 13:57:59  
Job time : 46.6467 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:20:38 ; Search time 142.551 Seconds  
(without alignments)  
2169.009 Million cell updates/sec

Title: US-09-917-376-3  
Perfect score: 4036  
Sequence: 1 ATTQPTWSNVAIGGGFVD.....YIGNGRGIIVGDIIGGAFSG 740

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US108\_PUBCOMB.pap.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | ID                   | Description       |
|------------|--------|-------------|--------|----------------------|-------------------|
| 1          | 4036   | 100.0       | 740    | US-09-917-376-3      | Sequence 3, Appli |
| 2          | 4036   | 100.0       | 740    | US-09-917-376-6      | Sequence 6, Appli |
| 3          | 4036   | 100.0       | 740    | US-10-155-400-3      | Sequence 3, Appli |
| 4          | 4036   | 100.0       | 740    | US-10-155-400-6      | Sequence 6, Appli |
| 5          | 4036   | 100.0       | 957    | US-09-917-376-1      | Sequence 1, Appli |
| 6          | 4036   | 100.0       | 957    | US-10-155-400-1      | Sequence 1, Appli |
| 7          | 2429.5 | 60.2        | 882    | US-10-156-761-9395   | Sequence 9395, Ap |
| 8          | 1680   | 41.6        | 726    | US-09-917-376-7      | Sequence 7, Appli |
| 9          | 1680   | 41.6        | 726    | US-10-155-400-7      | Sequence 7, Appli |
| 10         | 1579   | 39.1        | 838    | US-10-420-191-2      | Sequence 2, Appli |
| 11         | 1573.5 | 39.0        | 818    | US-10-026-994-2      | Sequence 2, Appli |
| 12         | 1405   | 34.8        | 739    | US-10-156-761-10111  | Sequence 10111, A |
| 13         | 1117   | 27.7        | 789    | US-10-395-241-14     | Sequence 14, Appl |
| 14         | 1117   | 27.7        | 812    | US-10-395-241-12     | Sequence 12, Appl |
| 15         | 1114   | 27.6        | 826    | US-10-395-241-18     | Sequence 18, Appl |
| 16         | 841.5  | 20.8        | 555    | US-09-927-827-47     | Sequence 47, Appl |
| 17         | 238    | 5.9         | 2468   | US-10-246-330-4      | Sequence 4, Appli |
| 18         | 238    | 5.9         | 2468   | US-10-282-122A-66335 | Sequence 66335, A |
| 19         | 199.5  | 4.9         | 2435   | US-10-282-122A-47453 | Sequence 47453, A |
| 20         | 193    | 4.8         | 1465   | US-10-369-493-13955  | Sequence 13955, A |
| 21         | 181    | 4.5         | 1308   | US-10-369-493-44566  | Sequence 44566, A |
| 22         | 172.5  | 4.3         | 783    | US-10-369-493-14105  | Sequence 14105, A |
| 23         | 170    | 4.2         | 1074   | US-10-282-122A-50616 | Sequence 50616, A |
| 24         | 168.5  | 4.2         | 1439   | US-10-282-122A-44567 | Sequence 44567, A |
| 25         | 167.5  | 4.2         | 1119   | US-10-245-802-12     | Sequence 12, Appl |
| 26         | 165.5  | 4.1         | 613    | US-10-369-493-11181  | Sequence 11181, A |
| 27         | 163.5  | 4.1         | 1289   | US-10-282-122A-44901 | Sequence 44901, A |

ALIGNMENTS

RESULT 1

US-09-917-376-3  
; Sequence 3, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Catalytic domain GH74  
US-09-917-376-3

| Query Match | Best Local Similarity | Score  | DB 3; | Length     | 740; |
|-------------|-----------------------|--|-------|------------|------|
| Matches     | 740;                  | Conservative   | 0;    | Mismatches | 0;   |
| Indels      | 0;                    | Gaps   | 0;    |            |      |
| Qy          | 1                     | ATTQPTWSNVAIGGGFVDGIVFNEGAGPGLIYVRDIDGMYRDWDAANGRWIPLLDVWGV  | 60    |            |      |
| Db          | 1                     | ATTQPTWSNVAIGGGFVDGIVFNEGAGPGLIYVRDIDGMYRDWDAANGRWIPLLDVWGV  | 60    |            |      |
| Qy          | 61                    | NMWGVVSVIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG       | 120   |            |      |
| Db          | 61                    | NMWGVVSVIADPINTNKVAAVGMVYNSWDPNDGAILRSSDQATWQITPLPKLGG       | 120   |            |      |
| Qy          | 121                   | NMPGRCGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNPPDVGTYIANPTDT   | 180   |            |      |
| Db          | 121                   | NMPGRCGERLAVDPNNNDILYFGAPSGKGLWSTDSGATWSQMTNPPDVGTYIANPTDT   | 180   |            |      |
| Qy          | 181                   | TYQSDIQGVVWVAFDKSSSLQASKTIFFGVADPNNPVFVSRDGGATWQAVPGATGF     | 240   |            |      |
| Db          | 181                   | TYQSDIQGVVWVAFDKSSSLQASKTIFFGVADPNNPVFVSRDGGATWQAVPGATGF     | 240   |            |      |
| Qy          | 241                   | IPHKGVDPNVHLYIATSNTCGPDYDGGSDGVWKFVTSCTWTRISPVSTDTANDYFGY    | 300   |            |      |
| Db          | 241                   | IPHKGVDPNVHLYIATSNTCGPDYDGGSDGVWKFVTSCTWTRISPVSTDTANDYFGY    | 300   |            |      |
| Qy          | 301                   | SGLTIDRQHPNTIMVATQISWNPDTTIFRSTDCGATWTRIDWTSYPNRSIRLVYLDISAE | 360   |            |      |



Db 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 360  
 Qy 361 PMLTFGVQPNPPVPSKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
 Db 361 PMLTFGVQPNPPVPSKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
 Qy 421 APWKGLLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTTGTSDVY 480  
 Db 421 APWKGLLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTTGTSDVY 480  
 Qy 481 AELNPSIIVRAGSPDPSSQPNDRHVAIFSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAIFSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTDDGVTF 600  
 Db 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTDDGVTF 600  
 Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRV 720  
 Db 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRV 720  
 Qy 721 YIGTNGRGIVYGDIGGAPSG 740  
 Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 2

US-09-917-376-6  
 ; Sequence 6, Application US/09917376  
 ; Publication No. US20040038334A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOLYTICUS  
 ; FILE REFERENCE: 40197.4US01  
 ; CURRENT APPLICATION NUMBER: US/09/917,376  
 ; CURRENT FILING DATE: 2001-07-28  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 US-09-917-376-6

Query Match 100.0%; Score 4036; DB 3; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-284;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ATTPYTWNSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 Db 1 ATTPYTWNSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 Qy 61 NNWGYNGVSIADDPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPKLGG 120  
 Db 61 NNWGYNGVSIADDPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPKLGG 120  
 Qy 121 NMPGRGMRGLAVDPNNNDNIIYFCAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
 Db 121 NMPGRGMRGLAVDPNNNDNIIYFCAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDT 180  
 Qy 181 TGQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWSRDGGATWQAVPGATPF 240  
 Db 181 TGQSDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPFWSRDGGATWQAVPGATPF 240

Qy 241 IPHKGVDPVNHVLYIATSNTCGPYDGSDDYWKFSVTSCTWTRISPVSTDTANDYFGY 300  
 Db 241 IPHKGVDPVNHVLYIATSNTCGPYDGSDDYWKFSVTSCTWTRISPVSTDTANDYFGY 300  
 Qy 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 360  
 Db 301 SGLTIDRQHPNTIMVATQISWMPDIIIFRSTDDGATWTRIDWTSYPNRSRLRYVLDISAE 360  
 Qy 361 PMLTFGVQPNPPVPSKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
 Db 361 PMLTFGVQPNPPVPSKLGWDEAMAIIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
 Qy 421 APWKGLLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTTGTSDVY 480  
 Db 421 APWKGLLEETAVNDLISPPSGAPLISALGDLGGTHADVTAVPSTIIFTSVPVFTTGTSDVY 480  
 Qy 481 AELNPSIIVRAGSPDPSSQPNDRHVAIFSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 540  
 Db 481 AELNPSIIVRAGSPDPSSQPNDRHVAIFSTDDGKNWFGQSEPGVTTGGTVAASADGSRFV 540  
 Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTDDGVTF 600  
 Db 541 WAPGDPQPVVYAVFGNSWAASQGVPANAIQIRSDRVNPKTFYALSNGTFYRSTDDGVTF 600  
 Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Db 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660  
 Qy 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRV 720  
 Db 661 KSAFGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDOHQYGNWGOAITGDHANLRV 720  
 Qy 721 YIGTNGRGIVYGDIGGAPSG 740  
 Db 721 YIGTNGRGIVYGDIGGAPSG 740

RESULT 3

US-10-155-400-3  
 ; Sequence 3, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOLYTICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 740  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE: Catalytic domain GH74  
 ; OTHER INFORMATION: Catalytic domain GH74  
 US-10-155-400-3

Query Match 100.0%; Score 4036; DB 4; Length 740;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-284;  
 Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 ATTPYTWNSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 Db 1 ATTPYTWNSNVAIGGGFVDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
 Qy 61 NNWGYNGVSIADDPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPKLGG 120  
 Db 61 NNWGYNGVSIADDPINTNKVAAVGMVTSNDPNDGAILRSSDQATWQITPLPKLGG 120

QY 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDCKSSSLGQASKTI FVGADPNPNPFWMSRDGGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDCKSSSLGQASKTI FVGADPNPNPFWMSRDGGATWQAVPGAPTGF 240  
QY 241 IPHKGVDPVNVHLVYIATNTGGPYDGSGLQWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 241 IPHKGVDPVNVHLVYIATNTGGPYDGSGLQWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 301 SGLTIDROHPNTIMVATQISWMPDTIIFRSTGGATWTRIMDWTSPNRSRYVLDISAE 360  
DB 301 SGLTIDROHPNTIMVATQISWMPDTIIFRSTGGATWTRIMDWTSPNRSRYVLDISAE 360  
QY 361 PMLTFGVQPNPVPSPKLGWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PMLTFGVQPNPVPSPKLGWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDY 480  
DB 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
DB 541 WAPGDPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVMFHAPCKEGDLWLAASSGLYHSTNGSSWSAITGYSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGVMFHAPCKEGDLWLAASSGLYHSTNGSSWSAITGYSSAVNVGFG 660  
QY 661 KSAPEGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
DB 661 KSAPEGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720

## RESULT 4

US-10-155-400-6  
; Sequence 6, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 740  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosyticus  
US-10-155-400-6

Query Match 100.0%; Score 4036; DB 4; Length 740;  
Best Local Similarity 100.0%; Pred. No. 4,7e-284;  
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTPQYTWNSVAIGGGFVDGIVFNEGAPGLIYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
|||||

DB 1 ATTPQYTWNSVAIGGGFVDGIVFNEGAPGLIYVRTDIGMYRWDAAANGRWIPLLDWVGW 60  
QY 61 NNWNGVGVVSIADPINTNKVAAVGMVYNSWPNPDGAILRSSDQAGATWQITPLPFLKGG 120  
DB 61 NNWNGVGVVSIADPINTNKVAAVGMVYNSWPNPDGAILRSSDQAGATWQITPLPFLKGG 120  
QY 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
DB 121 NMPGRGMRERLAVDPNNNDILYFGAPSGKGLWRSTDGATWSQMTNFPDVGTYIANPTDT 180  
QY 181 TGYQSDIQGVVWVAFDCKSSSLGQASKTI FVGADPNPNPFWMSRDGGATWQAVPGAPTGF 240  
DB 181 TGYQSDIQGVVWVAFDCKSSSLGQASKTI FVGADPNPNPFWMSRDGGATWQAVPGAPTGF 240  
QY 241 IPHKGVDPVNVHLVYIATNTGGPYDGSGLQWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 241 IPHKGVDPVNVHLVYIATNTGGPYDGSGLQWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 301 SGLTIDROHPNTIMVATQISWMPDTIIFRSTGGATWTRIMDWTSPNRSRYVLDISAE 360  
DB 301 SGLTIDROHPNTIMVATQISWMPDTIIFRSTGGATWTRIMDWTSPNRSRYVLDISAE 360  
QY 361 PMLTFGVQPNPVPSPKLGWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
DB 361 PMLTFGVQPNPVPSPKLGWDEAMAI DPNSDRMLYGTGATLYATNDLTKWDSGGQIHI 420  
QY 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDY 480  
DB 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSVPFTTGTSDY 480  
QY 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540  
DB 481 AELNPSIIVRAGSDPSSQPNDRHVAFTDGGKWFQSGSEPGVTTGGTVAASADGSRFV 540  
QY 541 WAPGDPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
DB 541 WAPGDPQPVVYAVGFGNSWAASQGVPAQAQIRSDRVNPKTFYALSNGTFYRSTDGGVTF 600  
QY 601 QPVAAGLPSSGAVGVMFHAPCKEGDLWLAASSGLYHSTNGSSWSAITGYSSAVNVGFG 660  
DB 601 QPVAAGLPSSGAVGVMFHAPCKEGDLWLAASSGLYHSTNGSSWSAITGYSSAVNVGFG 660  
QY 661 KSAPEGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
DB 661 KSAPEGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRV 720  
QY 721 YIGTNGRGIVYGDIGGAPSG 740  
DB 721 YIGTNGRGIVYGDIGGAPSG 740

## RESULT 5

US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES

; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-09-917-376-1

```
Query Match      100.0%; Score 4036; DB 3; Length 957;
Best Local Similarity 100.0%; Pred. No. 6.7e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 47 ATTQPTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106

Qy 61 NNWGYNGVVSIAADPINTNKVAAVGMVYTNWDNDGAILRSSDQGATWQITPLPFKLG 120
Db 107 NNWGYNGVVSIAADPINTNKVAAVGMVYTNWDNDGAILRSSDQGATWQITPLPFKLG 166

Qy 121 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYTIANPTDT 180
Db 167 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYTIANPTDT 226

Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240
Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286

Qy 241 IPHKGVPDPVNHVLIYATSNITGPGYDSSGVDWKFVSFTSGTWTTRISVPVSTDTANDYFGY 300
Db 287 IPHKGVPDPVNHVLIYATSNITGPGYDSSGVDWKFVSFTSGTWTTRISVPVSTDTANDYFGY 346

Qy 301 SGLTIDRQHPTNIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSRYVLDISAE 360
Db 347 SGLTIDRQHPTNIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSRYVLDISAE 406

Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420
Db 407 PWLTFGVQPNPVPSPKLGWDEAMADPPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 466

Qy 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480
Db 467 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 526

Qy 481 AELNPSIIVRAGSPDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540
Db 527 AELNPSIIVRAGSPDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 586

Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFFRSTDDGVTFF 600
Db 587 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFFRSTDDGVTFF 646

Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
Db 647 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706

Qy 661 KSAPGSSYPVAVVGTGGTGTGAYRSDCCGTWVLINDDQHOYGNWGOAITGDHANLRV 720
Db 707 KSAPGSSYPVAVVGTGGTGTGAYRSDCCGTWVLINDDQHOYGNWGOAITGDHANLRV 766

Qy 721 YIGTNGRGIVYGDIGGAPSG 740
Db 767 YIGTNGRGIVYGDIGGAPSG 786
```

## RESULT 6

US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; FILE REFERENCE: NREL 01-36A

; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

```
Query Match      100.0%; Score 4036; DB 4; Length 957;
Best Local Similarity 100.0%; Pred. No. 6.7e-284;
Matches 740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTQPTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 60
Db 47 ATTQPTWSNVAIGGGGFDGIVFNEGAPGILYVRTDIGMYRWDAAANGRWIPLLDWVGW 106

Qy 61 NNWGYNGVVSIAADPINTNKVAAVGMVYTNWDNDGAILRSSDQGATWQITPLPFKLG 120
Db 107 NNWGYNGVVSIAADPINTNKVAAVGMVYTNWDNDGAILRSSDQGATWQITPLPFKLG 166

Qy 121 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYTIANPTDT 180
Db 167 NMPGRGGERLAVDPNNDNIIYFGAPSGKGLWRSTDGATWSQMTNPDVGYTIANPTDT 226

Qy 181 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 240
Db 227 TGYQSDIQGVVWVAFDKSSSLGQASKTIIFGVADPNPNPVFWSRDGGATWQAVPGAPTGF 286

Qy 241 IPHKGVPDPVNHVLIYATSNITGPGYDSSGVDWKFVSFTSGTWTTRISVPVSTDTANDYFGY 300
Db 287 IPHKGVPDPVNHVLIYATSNITGPGYDSSGVDWKFVSFTSGTWTTRISVPVSTDTANDYFGY 346

Qy 301 SGLTIDRQHPTNIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSRYVLDISAE 360
Db 347 SGLTIDRQHPTNIMVATQISWMPDTIIFRSTDDGATWTRIDWTSYPNRSRYVLDISAE 406

Qy 361 PWLTFGVQPNPVPSPKLGWDEAMADPPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 420
Db 407 PWLTFGVQPNPVPSPKLGWDEAMADPPFNSDRMLYGTGATLYATNLDLTKWDSGGQIHI 466

Qy 421 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 480
Db 467 APVKGLEETAANDLISPPSGAPLISALGDLGGFTHADVTAVPSTIIFTSVFTTGTSDY 526

Qy 481 AELNPSIIVRAGSPDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 540
Db 527 AELNPSIIVRAGSPDPSSQPNDRHVAFTDGGKWFQSGEPGGVTTGGTVAASADGSRFV 586

Qy 541 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFFRSTDDGVTFF 600
Db 587 WAPGDPQPVVYAVFGNSWAASQGVPANAOIRSDRVNPKTFYALSNGTFFRSTDDGVTFF 646

Qy 601 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 660
Db 647 QPVAAGLPSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGGSSWSAITGVSSAVNVGFG 706

Qy 661 KSAPGSSYPVAVVGTGGTGTGAYRSDCCGTWVLINDDQHOYGNWGOAITGDHANLRV 720
Db 707 KSAPGSSYPVAVVGTGGTGTGAYRSDCCGTWVLINDDQHOYGNWGOAITGDHANLRV 766

Qy 721 YIGTNGRGIVYGDIGGAPSG 740
Db 767 YIGTNGRGIVYGDIGGAPSG 786
```

## RESULT 7

US-10-156-761-9395



Db 524 DGTVLLMSSTSGALVSKSQG--TLTAVSSLPSCGAVIASDKSDNTVYFGSAGAIYVSK 580  
 Qy 595 DGGVTFQFVAAAGLSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAI-TGVSS 653  
 Db 581 NTATSFKTIVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTA 638  
 Qy 654 ANNVGFGKAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQOA--IT 711  
 Db 639 GWSFGFGKASSTGSYVVIYGFFTIDGAAGLFSKSDAGTNMQVSDASHGFGS-GSANVYN 697  
 Qy 712 GDHANLRVYIGTNGRGIVYGDIGCAPSG 740  
 Db 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726  
 RESULT 9  
 US-10-155-400-7  
 ; Sequence 7, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; FILE REFERENCE: CELULOLYTICUS  
 ; CURRENT APPLICATION NUMBER: US/01-36A  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 7  
 ; LENGTH: 726  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus aculeatus  
 US-10-155-400-7

Query Match 41.6%; Score 1680; DB 4; Length 726;  
 Best Local Similarity 46.2%; Pred. No. 4.7e-113;  
 Matches 346; Conservative 113; Mismatches 258; Indels 32; Gaps 17;  
 Qy 1 ATTQPYTWSNAI-GGGGFGVGVNFGAGILYVTRTDIGMYRWDANGRWIPLLDVWG 59  
 Db 1 AASQAYTKNVVTVGGGGGFTGPIVFNPSAKGVAVARTDIGGAYRLN-SDDTWTPLMDVWG 59  
 Qy 60 ---WNWNGYGVSIADPINTNKVAAVGMVNTSWDPNDGAILRSSDQATWQITPLPP 116  
 Db 60 NDTWHDW---GIDALATDPVTDTRVYVAVGMVNTSWDPNDGAILRSSDQATWQITPLPP 116  
 Qy 117 KLGNNPGRGWERLAVDPNNNDNLYFGAPSGKGLNRSTDSGATWSQMTNFPDVGTYIAN 176  
 Db 117 KVGNNPGRGWERLAVDPNNNDNLYFGAPSGKGLNRSTDSGATWSQMTNFPDVGTYIAN 176  
 Qy 177 PTDITGQSDIQGVVVAFAFKSSSLGQASKTIFGVADPNPNVFWRSRGGATWQAVPGA 236  
 Db 177 SSST--YTSDPVLAWITFDSTSGSSGATPRIFGVADAGKSVKSEDAGATWAVSGE 234  
 Qy 237 PT-GFPHKGVDPVNVHLVIASTNGPVGSGGVDWKFVSVTGWTTRISVPVSTDTAN 295  
 Db 235 PQYGFGLPHKGVLSPEEKTLYISYANGAGPYDGTNGTVHKNIITSGVWTDISP---TSLAS 291  
 Qy 296 DYFGYGLTIDRHPNTIMVATQISWPDITIIIRSTDGGATWTRIMDWTSPNRSRYVL 355  
 Db 292 TYTYGGLSVLDLQVPGTFLMVAALNCWPPDELIIRSTDSGATWSPIWNGYPSINYYS 351  
 Qy 356 DISAEPWLTFGVQPNP-PVPSPKLGMDMAIDPNSDRMLYGTGATLYATNDLTKWDS 414  
 Db 352 DISNAPWIDTSTDQFPV---RVGMVVAALADPDSNHLVYGTGLTVYGGHDLTWNDS 408  
 Qy 415 GGOIHIAPMVKLEETAVNDLISPPSGAPLISALGDLGGFTHADVAVPSTIFTSVFTT 474  
 Db 409 KHNVTVKSLAVTEEMAVLGLITPPGFPALLSAVDDGGFVHSDLAAPNQAYHTPYGT 468

Qy 475 GTSVDYAEINPSIIVRAGSFPDSSQPNDRHVAFTSTGGKKNWFGSGEPGGVTTGCTGAASA 534  
 Db 469 TNGDITAGNKPNSIVRSASD--DYPT---LALSSNFGSTWADYAASTSTGTGAVALSA 523  
 Qy 535 DGSFVWAPGDPGPVVVAVGVGNSMAASQGVPAQAQIESDRVNPKTFFYALNNGTFFVRS 594  
 Db 524 DGTVLLMSSTSGALVSKSQG---TLTAVSSLPSCGAVIASDKSDNTVYFGSAGAIYVSK 580  
 Qy 595 DGGVTFQFVAAAGLSSGAVGMFHAVPGKEGDLWLAASSGLYHSTNGSSWSAI-TGVSS 653  
 Db 581 NTATSFKTIVS-LGSSTTVNAI-RAHPSIAGDVWASTDKGLWHSTDYGSTFTQIGSGVTA 638  
 Qy 654 ANNVGFGKAPGSSYPVAVVGTIGGVTGAYRSDDCGTTWVLINDDQHOYGNWQOA--IT 711  
 Db 639 GWSFGFGKASSTGSYVVIYGFFTIDGAAGLFSKSDAGTNMQVSDASHGFGS-GSANVYN 697  
 Qy 712 GDHANLRVYIGTNGRGIVYGDIGCAPSG 740  
 Db 698 GDLQTYGRVFRGHERPGHLLRQSQREPAG 726  
 RESULT 10  
 US-10-420-191-2  
 ; Sequence 2, Application US/10420191  
 ; Publication No. US20040067569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
 ; APPLICANT: Rey, Michael W.  
 ; APPLICANT: Zaretsky, Elizabeth J.  
 ; APPLICANT: Haas, Jeffrey A.  
 ; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
 ; FILE REFERENCE: 10210-200-US  
 ; CURRENT APPLICATION NUMBER: US/10/420,191  
 ; CURRENT FILING DATE: 2003-04-18  
 ; PRIOR APPLICATION NUMBER: US 60/373,987  
 ; PRIOR FILING DATE: 2002-04-19  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 2  
 ; LENGTH: 838  
 ; TYPE: PRT  
 ; ORGANISM: Trichoderma reesei  
 US-10-420-191-2

Query Match 39.1%; Score 1579; DB 4; Length 838;  
 Best Local Similarity 43.0%; Pred. No. 1.2e-105;  
 Matches 325; Conservative 123; Mismatches 264; Indels 44; Gaps 20;  
 Qy 6 YTWNSVAI-GGGGFGVGVNFGAGILYVTRTDIGMYRWDANGRWIPLLDVW---GW 60  
 Db 21 FSKRNKVLGGGGGVPVGIIFHPKTKGVAYARTDIGGLYRLN-ADDSWTAVTGTADIADNAG 79  
 Qy 61 NNWGYGVVSIADPINTNKVAAVGMVNTSWDPNDGAILRSSDQATWQITPLPPKLG 120  
 Db 80 HNW---GIDAVALLPQDDQKVYAAVGMVNTSWDPNSGAILRSSDRGATWSFNLPPKVG 136  
 Qy 121 NMPGRGWERLAVDPNNNDNLYFGAPSGKGLNRSTDSGATWSQMTNFPDVGTYIANPTD 180  
 Db 137 NMPGRGWERLAVDPNSNIIYFGARSGNGLWKS TDGGVTFKSVSFTATGTYIPDPSDS 196  
 Qy 181 TGYQSDTQGVVWVAFDKSSSLGQASKTIFGVADPN-NNPVFWRSRGGATWQAVPGAPT 239  
 Db 197 NGYNSDKQLMMVTFDSTSTTGGATSRIFVGTADNITASVYVSTNAGTSVAVPQPGK 256  
 Qy 240 FIPKGVDPVNVHLYIATSNITGPGYDGSQGVWKFVSVTGWTTRISVPVSTDTANDYFG 299  
 Db 257 YFPHAKLQPAEKALYLYTSGTGYDGTGLSVWRIDYTAGGWTGKDIITVSGSDL---YFG 313  
 Qy 300 YSLGTLDRQHENTIMVATQISWPDITIIIRSTDGGATWTRIMDWTSPNRSRYVLDISA 359  
 Db 314 FGLGLDLQKQGLTAVASLNSWMPDQALFRSTDSCTTWSPIWAWASYPTETYYISISTPK 373



QY 126 GMERLAVDPNNNLLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQS 185  
 DB 161 GAGERLLVDPDSDTLWLGT-RHDGLLKSTDRGATMAAATAFP-----AKANS 207  
 QY 186 DIOGVVAVFADKSSSSISGQASKTIFVGVADPNP-----VWSDRGATQAQVGPATGF 240  
 DB 208 SGOGVVVF-----LVAAGRTVYAGWGDGDTSGTANLYRTAD-GTTWGAVGPSPSGT 257  
 QY 241 ---IPHKGVDPVNVHLYIATSTNGPYDGSDDVWKFSTVGTWTRISPV---PSTDT 293  
 DB 258 SAKVPLRAAYDTHRELYVYTGAPGCGSDSVHKLRTATGTWTEVTPVKPGCTTSDG 317  
 QY 294 ANDYFGSLTIDROHNTIMVATQISWMPDTIIFRSTDSGATWTRIDWTSYFNRLRY 353  
 DB 318 SADTFAYGGVAVDARPGTLVSTNNRWADGDTVFRSTDSGRTWTSKLD-----AA 368  
 QY 354 VLDISAPMLTFVQVQNPVPSKLGWDEAMALDPNSDRMLYGTGATLYATNDLTWMD 413  
 DB 369 VFDVSETPFLDWGDD-----KPKFGWMIQALAVDPYDSQVYVYGTGATLYGTRDLKRW- 421  
 QY 414 SGQIHLAPMWKLEETAVNDLISPSG-APLISALGDLGFTHADVTAVPST-IETSPV 471  
 DB 422 -----APRILEESAVNQLISPPVGEAHLISGLDIGNVHERLUTASPSGMAATNPV 474  
 QY 472 FITTGSVDVAELNPSIIVRAGSPDPSPQPNDRHVAFSTDSGKNW---PQSEPGGVTTGGT 529  
 DB 475 FGSATGLAQAAAPAYVVRTGWGDHNG-----AYSHDGRTWAPFEAQDIADKADGP 528  
 QY 530 VAASAGSRFVA-----PGDGPVVVA-----VGFNGSWAASQGVPAQAQIRSDRVNPKT 581  
 DB 529 IATSDAGGTLMSFVHWDG-----TTVAHRSTDNGASWSEVSFPKGAFTVADPADPTR 583  
 QY 582 FYA--LSNGTFYESTDGGVTFQFVAAGLPSSGAVGVMFHAPVCKEGDLWLAAS-SGLYHS 638  
 DB 584 FYAYDFNGLYASTDSGRSFTARAGLP-SGDSQFKLVAAPGRSGDLWLSAKWNGLYRS 642  
 QY 639 TNGSSSAITGVSSAVNVGFGKSPAGSSYPVAVFVGTIGGVTCAYRSDDCGTTWVLND 698  
 DB 643 TDGDTFARIDSCWASYTLGFGAAGADYPAIYQVGSSTETITAVYRSDDAARTWRLND 702  
 QY 699 DOHQYNGWQAIGTDHANLRVYVGTNGRGIVYGD 733  
 DB 703 DAHOWGIGEAUVGDPRHGRVYLATNGRGIQYE 737

RESULT 13  
 US-10-395-241-14  
 ; Sequence 14, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; FILE REFERENCE: 073756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 14  
 ; LENGTH: 789  
 ; TYPE: PRT  
 ; ORGANISM: Geotrichum sp. M128  
 US-10-395-241-14

Query Match 27.7%; Score 1117; DB 4; Length 789;  
 Best Local Similarity 35.0%; Pred. No. 3.9e-72;  
 Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;  
 QY 6 YTWNSVAIGGGFVDGIVFNEGAPGILYVTRDITGGMYRWDAAANGRWIPLLDWYGNWGWY 65

DB 4 YEFKNVAIGGGYITGIVAHPKTKDLYARTDIGAYRWDAGTSKWIPLDNFIADQMN 13  
 QY 66 NGVYSIAADPINTKNVAAVAGMYT-NSWDPNDAIGALLRSSDQCATWQITPLPKLGNMGP 124  
 DB 64 MGTESIALDPNPRRLYLAQRYGVDEW-----AAYVEDRQGSFTIYESPPMGANDMG 119  
 QY 125 RGMERLAVDPNNNLLYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTGYQ 184  
 DB 120 RNNGERLAVNPFNSNEVWMTGRT-EGIKSSDRAKTWTNVTSPDAF-----TNGIGYT 172  
 QY 185 SDIQGVVWVADPKSSSSISGQASKTIFVGVADPNPVPFWSRDCGATWQAVPGAPTGF 241  
 DB 173 S-----VIDP-----ERNGTIYASATAPQG-MYVTHDGVSWEPVAGQPSMLNRT 218  
 QY 242 -----PHKGVDPVNVHLYIATSTNGPYDGSDDVWKFSTVGTWTRISPV---PSTDT 293  
 DB 219 TGAPDKKPAIAPOPMKVALTP--NFLYTVADYPPGPGWVTFGEVWRQNTSGAWDDIT 276  
 QY 286 -----SPVSTDTANDYFGYSLTIDROHNTIMVATQISWMPDTI---IFRSTDSGAT 336  
 DB 277 PRVGNSSPAPYNNQTFPAGGFCGLSDATNPNLVIT-LDRDPGALDSIYLSTADAGT 335  
 QY 337 WTRIWDWTS-----YPNRSLRYVLDISAEPMLTFGVQP-----NPPVPSP---KLG 379  
 DB 336 WKDVTQLSSPSNLEGNWGHPTAARY-KDGTVPVWLDNFNNGPQWGGYGAPHGTPGLTKFG 394  
 QY 380 WMDEAMALDPNSDRMLYGTGATLYATNDLTG-----WDSGGQIHAP-----MVKGLEETA 431  
 DB 395 WMSAVLIDPFPNPHLMYGTGATWATDLSRVEKW-----APSWYLOIDGIEENA 446  
 QY 432 VNDLISPPSAPLISALGDLGFTHADVTAVPSTIFTSPVFTTGTSDVVAELNPSIIVRA 491  
 DB 447 ILSRSPKSGAALLSGIDISGMKEDDLTK-PQKMGAPQFNSNLSIDAGNFPNVVRA 505  
 QY 492 GSPDPSPQPNDRHVAFSTDSGKNW---FQSEBPGGVTT---GGTVAASADGSRFVWAP--G 544  
 DB 506 GSSGHEYDSACARGAYATDGDWATIFPTCPGNWASHVQGSTIAVDASGQIWSKLD 565  
 QY 545 DPGQPVVYAVFGNSWAAASQVPA-----NAQTRSDRVNPKTFYALSNGTYRSTDSGV 598  
 DB 566 EQASGPWYSHDYGKTWS-----VPAGDLKAQATANVLSKVDQGTFTYATDGGKFFVSTGGK 621  
 QY 599 TFQPVAAAGLPSSGAVGVMFHAPVCKEGDLWL-AAASSGLYHSTNGGSSKSAI-TGVSSAVN 656  
 DB 622 SYAAGKAGLV--GTSLMPAVNVPVAGVWVPEGGFLHSTDFGASFTRVGTANATLVS 679  
 QY 657 VGFGKS-----APGSSYPVAVFVVGVT--IGGVTCAYRSDDCGTTWVLINDDQHOYNGWQAI 710  
 DB 680 VGAPKSKDGKASAPSAVFWGTDKPGSDIGLYRSDDNGSTWTRVNDQEHYSG-PTWI 738  
 QY 711 TGDHANLRVYVGTNGRGIVYGD 734  
 DB 739 EADPKVYGRVYVLTNGRGIVYADL 762

RESULT 14  
 US-10-395-241-12  
 ; Sequence 12, Application US/10395241  
 ; Publication No. US20040038367A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAOI, Katsuro  
 ; APPLICANT: MITSUISHI, Yasushi  
 ; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
 ; FILE REFERENCE: 073756  
 ; CURRENT APPLICATION NUMBER: US/10/395,241  
 ; CURRENT FILING DATE: 2003-03-25  
 ; PRIOR APPLICATION NUMBER: JP 2002-83433  
 ; PRIOR FILING DATE: 2002-03-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 12



```
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Geotrichum sp. M128
US-10-395-241-12

Query Match      27.7%; Score 1117; DB 4; Length 812;
Best Local Similarity 35.0%; Pred. No. 4e-72;
Matches 281; Conservative 124; Mismatches 279; Indels 120; Gaps 31;

QY 6 YTNVAIGGGFVDGIVNEGAPGILYVRTDIGMYRMDAANGRWIPLLDWGNWNGY 65
Db 27 YEFKNVAIGGGYITGIVAHPKTDLLYARTDIGAYRWDAGTSKWIPLNDFIEAQDMNI 86
QY 66 NGVSIADDPINTNKVMAAVGMYT-NSWDPNDGAILRSSDQATWQITPLPKLGGNMPG 124
Db 87 MGTESIALDPNPDRLYLAAQRYVGDEW----AAFVSEDRGQSFTIYESPPPMGANDMG 142
QY 125 RGMGERLAVDPNNDNLYFGAPSGKGLWRSTDGATWSQMTNFPDVGVIYANPTDTTGYQ 184
Db 143 RNRGERLAVNPNFNSNEVWMTGRT-EGIKSSDRAKTWTNVTSPDAP-----TNGIGYT 195
QY 185 SDIQGVVWVAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDGGATWQAVPGAPTGF 241
Db 196 S-----VIFDP-----ERNGTIYASATAPOG-MYVTHDGGVSWEPVAGOPSSWLNR 241
QY 242 -----PHKGVDPVNHVLYIATSNTPGYPDSSGDVWKFVSVTSGTWTRI- 285
Db 242 TGAFFDKPKASTAPQPMKVALTP--NFLYVYADYFGPWGVTFGVWRQNRNRTSGAWDDIT 299
QY 286 -----SPVSTDTANDYFGYGLTIDRQHPNTIMVATQISWMPDTI---IFRSTDGAT 336
Db 300 PRVGNSSPAPYNNQTFPAGGFCGLSVDATNPNRLVIT-LDRDPGPALESLYSTDAGAT 358
QY 337 WTRTWDWTS-----YPNRSLRYVLDISAEPWLTFGVOP-----NPPVPSP---KLG 379
Db 359 WKDVTQLSSPNSLEGNWGHPTNAARY-KDGTVPVPLDNNPGQWGGYGAPHGTPGLTKFG 417
QY 380 WMDEAMIDPFNSDRMLYGTGATLYATNDLTK-----WDSGGQIHIAP-----MVKGLEETA 431
Db 418 WMSAVLIDPFNPEHLMYGTGATITWATDLSRVEKDW-----APSWYLQIDGIEENA 469
QY 432 VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVTTGTSTVDYAEINPSIIVRA 491
Db 470 ILSLRSPKSGAALLSGIGDISGMKHDDLTK-PQKMFAGAPQFSLNLSIDAAGNPNVAVRA 528
QY 492 GSFDPSSQPNDRHAFSTDGKNW--FQSGEPGGVTT---GGTVAASADGSRFWAP--G 544
Db 529 GSSGHEYDSACARGAYATDGGDAWTFPTCTPPGMNASHYQGSTIAVDASSQIWMSTKLD 588
QY 545 DPGQPVVYAVFGNSWAASQGVPA-----NAQIRSDRVNPKTFYALNNGTFYRSTDGGV 598
Db 589 EQASGPWYSHDYGKTWS-----VPAGDLKAQTANVLSDKVQGTFTYATDGGKFFVSTDGK 644
QY 599 TFQPVAAGLPSSGAVGMFHAVPGKEGDLWL-AAASSGLYHSTNGGSSWSAI-TGVSSAVN 656
Db 645 SYAAKAGAGLVT--GTSMPAVNPWVAGDVWVPVEGGLFHSITDFGASFTRVGTANATLVS 702
QY 657 VGFGKS-----APGSSYPAVFVGT--IGGVTAAYRSDDCGTTWVLINDDQHQYGNWQAI 710
Db 703 VGAPKSKSDGKKASAPSAVFITGTDKPGSDIGLYRSDDNGSTWTRVNDQEHNSG-PTMI 761
QY 711 TGDHANLRVYVIGTNGRGIVYGD 734
Db 762 EADPKYGRVYLTNGRGIVYADL 785

RESULT 15
US-10-395-241-18
; Sequence 18, Application US/10395241
; Publication No. US20040038367A1
; GENERAL INFORMATION:
; APPLICANT: YAOI, Katsuro
; APPLICANT: MITSUISHI, Yasushi
```

QY 711 TGDHANLRRVYIGTNGRGIVYVDI 734  
Db 740 EADPKVYGRVYLG TNGRGIVYADL 763

Search completed: March 2, 2006, 14:26:54  
Job time : 146.551 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:27:20 ; Search time 15.0053 Seconds  
(without alignments)  
986.322 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATQPTWNSNAIGGGGFVD.....YIGTNGRGIYVDIGGAPSG 740

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 135339 seqs, 20000136 residues

Total number of hits satisfying chosen parameters: 135339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID                  | Description        |
|------------|-------|-------------|--------|----|---------------------|--------------------|
| 1          | 175.5 | 4.3         | 2523   | 7  | US-11-052-554A-143  | Sequence 143, App  |
| 2          | 168.5 | 4.2         | 5291   | 7  | US-11-052-554A-281  | Sequence 281, App  |
| 3          | 166.5 | 4.1         | 980    | 7  | US-11-052-554A-17   | Sequence 17, App   |
| 4          | 160.5 | 4.0         | 3716   | 7  | US-11-052-554A-141  | Sequence 141, App  |
| 5          | 156.5 | 3.9         | 2204   | 7  | US-11-052-554A-134  | Sequence 134, App  |
| 6          | 156   | 3.9         | 3157   | 7  | US-11-052-554A-142  | Sequence 142, App  |
| 7          | 155   | 3.8         | 7465   | 7  | US-11-087-099-7521  | Sequence 7521, App |
| 8          | 153.5 | 3.8         | 807    | 7  | US-11-087-099-12161 | Sequence 12161, A  |
| 9          | 148.5 | 3.7         | 894    | 7  | US-11-087-099-11214 | Sequence 11214, A  |
| 10         | 147.5 | 3.7         | 755    | 6  | US-10-517-939-330   | Sequence 330, App  |
| 11         | 146   | 3.6         | 949    | 7  | US-11-052-554A-6    | Sequence 6, Appli  |
| 12         | 145.5 | 3.6         | 3300   | 7  | US-11-052-554A-133  | Sequence 133, App  |
| 13         | 145   | 3.6         | 760    | 7  | US-11-052-554A-68   | Sequence 68, Appl  |
| 14         | 140.5 | 3.5         | 955    | 7  | US-11-052-554A-179  | Sequence 179, App  |
| 15         | 140.5 | 3.5         | 1461   | 7  | US-11-052-554A-283  | Sequence 283, App  |
| 16         | 139   | 3.4         | 3063   | 7  | US-11-186-284-26    | Sequence 26, Appl  |
| 17         | 137.5 | 3.4         | 1436   | 7  | US-11-052-554A-140  | Sequence 140, App  |
| 18         | 137   | 3.4         | 1424   | 7  | US-11-102-476-4     | Sequence 4, Appli  |
| 19         | 136.5 | 3.4         | 393    | 5  | US-09-930-864-1     | Sequence 1, Appli  |
| 20         | 134.5 | 3.3         | 1007   | 6  | US-10-517-939-84    | Sequence 84, Appl  |
| 21         | 133.5 | 3.3         | 837    | 7  | US-11-052-554A-159  | Sequence 159, App  |
| 22         | 133.5 | 3.3         | 1889   | 7  | US-11-102-476-46    | Sequence 46, Appl  |
| 23         | 133.5 | 3.3         | 3132   | 7  | US-11-087-099-1245  | Sequence 1245, App |
| 24         | 130   | 3.2         | 1588   | 7  | US-11-052-554A-280  | Sequence 280, App  |
| 25         | 129.5 | 3.2         | 736    | 7  | US-11-085-185-4     | Sequence 4, Appli  |

|    |       |     |      |   |                     |                   |
|----|-------|-----|------|---|---------------------|-------------------|
| 26 | 129   | 3.2 | 1306 | 7 | US-11-052-554A-139  | Sequence 139, App |
| 27 | 126.5 | 3.1 | 778  | 7 | US-11-052-554A-144  | Sequence 144, App |
| 28 | 126   | 3.1 | 527  | 7 | US-10-886-504-10    | Sequence 10, Appl |
| 29 | 126   | 3.1 | 527  | 6 | US-10-886-505-10    | Sequence 10, Appl |
| 30 | 126   | 3.1 | 527  | 6 | US-10-886-527-10    | Sequence 10, Appl |
| 31 | 126   | 3.1 | 2105 | 7 | US-11-052-554A-173  | Sequence 173, App |
| 32 | 125.5 | 3.1 | 538  | 7 | US-11-174-398-16    | Sequence 16, Appl |
| 33 | 125.5 | 3.1 | 538  | 7 | US-11-132-947-6     | Sequence 6, Appli |
| 34 | 125.5 | 3.1 | 538  | 7 | US-11-197-488-2     | Sequence 2, Appli |
| 35 | 125   | 3.1 | 527  | 6 | US-10-886-504-8     | Sequence 8, Appli |
| 36 | 125   | 3.1 | 527  | 6 | US-10-886-505-8     | Sequence 8, Appli |
| 37 | 125   | 3.1 | 527  | 6 | US-10-886-527-8     | Sequence 8, Appli |
| 38 | 125   | 3.1 | 611  | 6 | US-10-517-939-156   | Sequence 156, App |
| 39 | 125   | 3.1 | 1417 | 7 | US-11-052-554A-8    | Sequence 8, Appli |
| 40 | 125   | 3.1 | 6893 | 7 | US-11-205-109-14    | Sequence 14, Appl |
| 41 | 124.5 | 3.1 | 461  | 7 | US-11-087-099-11993 | Sequence 11993, A |
| 42 | 124.5 | 3.1 | 1579 | 7 | US-11-052-554A-9    | Sequence 9, Appli |
| 43 | 124   | 3.1 | 525  | 7 | US-11-098-686-10993 | Sequence 10993, A |
| 44 | 124   | 3.1 | 957  | 7 | US-11-108-172-1065  | Sequence 1065, Ap |
| 45 | 122.5 | 3.0 | 863  | 7 | US-11-087-099-10485 | Sequence 10485, A |

#### ALIGNMENTS

#### RESULT 1

US-11-052-554A-143

; Sequence 143, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 143

; LENGTH: 2523

; TYPE: PRT

; ORGANISM: Mycobacterium tuberculosis H37Rv

US-11-052-554A-143

|                       |                  |   |             |              |
|-----------------------|------------------|---|-------------|--------------|
| Query Match           | 4.3%;            | Score 175.5;  | DB 7;       | Length 2523; |
| Best Local Similarity | 21.8%;           | Pred. No. 0.00041;  |             |              |
| Matches 175;          | Conservative 92; | Mismatches 327;   | Indels 209; | Gaps 38;     |
| QY                    | 10               | NVAICGGGFVDGIVFNEGA--PGILYVRTDIGMYRWDAAN-GRWIPLLDWGNNWYN      | 66          |              |
| DB                    | 1522             | NTGICNSGIASGTGLFNAGGFTGTVNAGSYNTGSPNAGQANTGGFPGSVNTCWLNTG--   | 1579        |              |
| QY                    | 67               | GVVSTAADPINTNKVAAVGMTNSWDNDGAILRSS--DQATWQITPLPKLGGNMPG--     | 124         |              |
| DB                    | 1580             | -----DINTG-----VANSQDVNTGAFISGNYNGAFWR-----GDYQGL             | 1614        |              |
| QY                    | 125              | RCMGERLAVDPNNDNILYFGAPSGKGLWRSTDGATWSQMTNPPDVGTIANPTDITGYQ    | 184         |              |
| DB                    | 1615             | LGFSYRPAVLQO-----TPFLDITLTTGGJGSGVVIPAIDPAIR                  | 1652        |              |
| QY                    | 185              | SDIQGVWVWV-AFDKSSSSLSGQ---ASKTIFVGVAD-----PNNPFWMSRDGGATWQA   | 232         |              |
| DB                    | 1653             | PEFSANVAIDSTVPSIPQIDLAATVSVGLGPITVPHLDIPRPVTLNLYFGSQ---       | 1709        |              |
| QY                    | 233              | VGAFTGPIPHKGFPDPVNNHLYIATSNVTGGPYDSSGSDVWKFSTVSGTWTRISPVST    | 292         |              |
| DB                    | 1710             | -PGGPKIGPITGLFNTPIGLTPLALSQIVIGASSSQGITITAFLANLPFTSTPVVTIDEIP | 1768        |              |
| QY                    | 293              | TANDYFGYS-----GLTIDRQHPNTI-----MVATQISWPDPIIFRSTDDG           | 334         |              |

Db 1769 LLASITGHSEPVDFPGGLIIPANNPLSINLSGCTGAVTIPAITIGIPDPLVAHSTLG- 1827  
Qy 335 ATWRIWMTSPYNSRLRYVLDISAEPMLTFGVQPNPVPSPKLGWDE- 383  
Db 1828 -----PVHILIDLPAVP--GFG-----NTTGAPSSGFFNSGAGGVSGFNV 1866  
Qy 384 -AMAIIDPFI-SDRMLYGTGATLYATNDLTKWDSGQIHIAPMWKGLBETAVNDLISPPSG 441  
Db 1867 GAVMSGWGNQAPSAALLGGSGVFNAGTL-----HSGVLNFGSGMGLFNTSVLGL-----G 1917  
Qy 442 AP-LISALGDLGGTHADVTAVPSTIFTPVFTTGTSDVDYAEALNPSIIVRAGSDPSSQP 500  
Db 1918 APALVSLGSGVQ-----QLSGLLASGTA-----LHQGLVLPGLADVGL-- 1957  
Qy 501 NDRHVAFTSDGGKMWFGSEBGGVTTGGTVAASADGSRFV-WAP-----GD-----PGQPVVY 552  
Db 1958 --GNVGLGNVDFNLGAGNVGPNVGGNIGNNVGLGNVGMGNFGLNSGLTGLMGLG 2015  
Qy 553 AVGFNSWAAASQGVPAANAQIRSDRVNPKTFVALSNGTF-----YRSTGGV 598  
Db 2016 NIGFNGAGSYNFGI-ANW-----GVGNIGFANTGSGNFGIGLTDNLGTGFGFNTSGNV 2069  
Qy 599 TFQPVAAGLPSSGAVGV-MHAPVKGEDLWLAASSGLYHSTNGSGSSWALTGVSSAVNV 657  
Db 2070 -----GLFNSGTVGNVGFNSGTGN-----WGVFNSSGY-NTGIGNSGIASGLFNAGGF 2117  
Qy 658 GFGKSPAGSSYPVAVVVTGGTGVAYRSDDCGTTWVLINDQHQYGNWQA-----ITG 712  
Db 2118 NTGVNAGSYNTGFSNAGQAN--TGGFNPVSGVNTGMLNTGINTGVANSQDVNTGATFISG 2175  
Qy 713 DHANLRVYITNGRGIVYGDIG 735  
Db 2176 NYSN-----GAFWRGDIYQGLLG 2192

## RESULT 2

US-11-052-554A-281  
; Sequence 281, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052.554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 281  
; LENGTH: 5291  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-281

Query Match 4.2%; Score 168.5; DB 7; Length 5291;  
Best Local Similarity 20.1%; Pred. No. 0.0031;  
Matches 190; Conservative 82; Mismatches 270; Indels 401; Gaps 46;

Qy 27 GARGILYVRDTCGMYRWDAANGRWIPLDWGNWNNYGVVSVIAADPINTKNVWAAVG 86  
Db 4186 GANFAQAISTDNGA-----TWNV-----VTVAADSLN----- 4212  
Qy 87 MYTNSDPNDGATILRSSDQATWQITLPLPKLGNMPCRGGERLAVDPNN----- 137  
Db 4213 -----WSYVDGRLTNGT--TTWQVRVV--DLAGNV--GATSSOSALIDTVNPAQLVTIAS 4262  
Qy 138 -----DNILYF-----GAPSGKGLMRSTDGATWSQMTNFPDVGTYIAN 176

Db 4263 ISTDGSSATDFITSDTMLITLGSILGASGEVAQISLDSGATWTTLTNGTQWTTYDS 4322  
Qy 177 PTDTTG-----YQSDIQGVVWVAFDSSSLGQASKTIFVGVADPN---NPVFWSRDGGATW 230  
Db 4323 RLTLDGSVYQVRV-----LDLAGNTGPVVKTKWVDVTINPTATPTIVSYTDDVGQRQ 4375  
Qy 231 QAVPGAPTGFPHKGVDFPDVNHVLYIATNSNTGPPYDG-----SSGDV----- 272  
Db 4376 GTLSSQ-----ATDDTTPLLNGVLASPLASGEVVLYRNGILLGA 4416  
Qy 273 -----WKFS-----RISPVSTDTANDYFGYSGLTIDRQHENTIMVAV- 317  
Db 4417 VTMVAGLNWYISDGLVSGAITYSARVVDLAGNITSSDF-----VLTVDSIPTTLAQITS 4473  
Qy 318 -----QISW-----PDT----- 325  
Db 4474 QTRDTTPIISGVITAALASGOYVEVINGKTYTSEPGGVAVVDPAHNTWYVQLPDTDAL 4533  
Qy 326 -----IIFRSTDDGATWTRIWDWTSYPNRSLRYVL 355  
Db 4534 TVSATAVTVAQVKSAGNGNNANISNGTVTVNAADYTPTWTTASKTTAW---GLTYGL 4590  
Qy 356 DISAEPMLTFGVQPNPVPSPKLGWDEAMALDPNSDRMLYGTGATLYATNDLTKWDSG 415  
Db 4591 D-SHGMVTVLANQOVMOSTDP-LTWSKTALT-----LYQSGNN-YATSSIAIDYDRN 4638  
Qy 416 QOIHIAPMWKLEETAVANDLISPPSGAPLIS-----ALGDLGFT-HADVAVPSTIFTSP 470  
Db 4639 G-----TGDLFITRDDYGTGYINGFTNGDGT-----FSSA 4669  
Qy 471 V-FTTGSVDYAEALNPSIIVRAGSDPSSQPNDRHAFST--DGKNNWFQSGEPG----- 522  
Db 4670 IQVTGVLTWYGS-----VAFDKEGDGYLDFEWIGDAGGPDSTNT 4708  
Qy 523 -----GVTTGCTVAASADGSRFVWAPGDPQVYVAVFGNSWAAASQGVPAANAQIRSDR 576  
Db 4709 FLWNAGTLVGNSTNSGGSATVGG-----AVTGYLSLNEGSGVDLNDGRIDL 4758  
Qy 577 V---NPKTFYALS-----NGTFVRSYTDGVTQFVAAAGLPSSGAGVGMFHAVPGK-EG 625  
Db 4759 VOHTYNLNYYTSLSLNQNGTTFWQONTNTFLSGAGSGAMSSSVSWTWADFDDGDM 4818  
Qy 626 DLWLAASSG-----LYHSTNG----- 641  
Db 4819 DLELPASQGRANYGSLLFNTNGVLGCPVAVGATATTAYASQFSLAVDWNHDLGLDIARIAQ 4878  
Qy 642 -GSSWSAITGVSSAVNVGFGKSPAGSSYPVAVVVTGIGVT-----GAYR---SDDCG 690  
Db 4879 TGQSY-LYTNVSNASN--WTQSALGSSQS-----GTTSGVAAMDYWDGAVDVLVSKQSG 4930  
Qy 691 TTWVLINDDQHQYGNWQAITGDHANLRVYVIGTNGRGIVYGD 733  
Db 4931 SVFLSRNTNTVSYG-----TSLHLRI-----TDPNGINVIYGN 4963

## RESULT 3

US-11-052-554A-17  
; Sequence 17, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052.554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 17



Db 1867 TSLGSLNAGTASGLNSCTAISGLFNVSTLDTTPAVISGFSNLGDMHSGV----SIDGL 1923  
Qy 598 V---TFQPV-----AAGLPSSGAGV-----MFHAVPGKEGD 626  
Db 1924 IAILTFPPAESVFDQIIIDAAELQHLIDGNALAGNVGGVNLGLANVGEENLGNAGNVN 1983  
Qy 627 LWLAASSGLYHSTNGGSSWAITGVSSAVNVGFGKAPGSSYPAVVVGTGTGGTVGAYRS 686  
Db 1984 INVAG-----NLGSGNLGLNVGTG-NLFGNIGAGN-----FGFAGLTAGAGGL 2030  
Qy 687 DDCGTTWVLNDQHQYGNWQCAITGDHANLRRVYVGTNGRGI-VYGD-----ICGAPSG 740  
Db 2031 GNVG-----LGNAGS-----GSMGLANVG-VGNGLANTGTGIGLTDYRTGIGELNSG 2081

RESULT 5  
US-11-052-554A-134  
; Sequence 134, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 134  
; LENGTH: 2204  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-134

Query Match 3.9%; Score 156.5; DB 7; Length 2204;  
Best Local Similarity 21.18; Pred. No. 0.0072;  
Matches 177; Conservative 91; Mismatches 287; Indels 283; Gaps 42;

Qy 9 SNVAIGGGFVDGIVFNEGAPGILYVRTDYGMYRWDAAAGRWIPLLDWVGWNNWG-YN- 66  
Db 100 NVVGIGNTGIFDVGLANLGSYNGF--GNLGD-----DNLGFGNFGSYNI 142  
Qy 67 GWSVIAADPIN-TNKVAAVGM-----YTNWDPNDAILRSSDQGTWQITPLPF 116  
Db 143 GFGNVGNDLGFANAGGNGIGFANTGNNVFGNGTGSNNVIGLGTNGQ----- 191  
Qy 117 KLGNNMFGRGWGERLAVDPNNDILYFGAPSGKGLWRSTDSGATWSQMTNPPDVGTVIAN 176  
Db 192 -IGFGNSGSGNIGLNSGNNIGFN--SGSNFGIANSFNTGIGNTNTGLFN 248  
Qy 177 PTD--TTCYQSDIQGVVWVAFDKSSSILGQASKTIF-VGVADPNPNPFWSRDGGATWQAVP 234  
Db 249 SGDVTG-----AFNPGSFNTGSGFNTGFGNPGN----- 281  
Qy 235 GAPTGFPHKGVPDPVNHVLIYATSNTPGPGYDGSQGVKFSVTSGTWTRISPVSPDTDA 294  
Db 282 -TNTGY-----LNIYNTG---IANTGVDGTGAFITGYSN-----GLFLS 319  
Qy 295 NDYFVSGLTIDRHPNTIWAQISWWPDTIIFRSTDDGATWTRIDWTSYPNRSRYV 354  
Db 320 GDYQGLVGLNLVIDMPLFISLGVNI---PIDIPITASAGNITLMGV---TIPPTGDI--V 371  
Qy 355 LDISAEPWLTFG--VQNPVPVSP-----KLGWDEMAIDPFNSDRMLYGTGATLYATND 408  
Db 372 LSSIAQORAHFGPTIIPNITVGGTTTVAIGGPNTAITI-----TG----- 412  
Qy 409 LTRWDSGGQIHI-----APMVKGLEETAVN-----DLIS 437

Db 413 -----CGAIRIPLISIPAAFGFGNSTTNPSSGFNTCAGGASGFGNFGANGSFGWNLAS 466  
Qy 438 PPSGAPLISALGDLG-----GFTHADVTAVPSTIFTSPVFTTGTSTV----- 478  
Db 467 ATSGASGLNLVAGLGLANVGTTVSGFNTSTSLATPAFNSGLANISTSIAGLLRDST 526  
Qy 479 -----DYAELNPSIIVRAGSFDPSQPNDRHAFSTGGKRW-----FQGSPPGV 524  
Db 527 GTMVLNLGLANHGTNLVGI-----ANLGDYNI GFANLGSANFGSANFGSANGNIGGA 576  
Qy 525 TTGGTVAASADGSFRVWAPGDPGPV-----YAVGFGNSWAASQGVPAQAIRSDRV 577  
Db 577 NTGIPDIGLANLGSYNGFGNFGDNDLGFNLGSYNGFGNLGNDNLGF-ANTG----- 629  
Qy 578 NPKTFYALNSGTFFRSTDCGVTFPVAAAGLPSSGAGVGMFHAVPGKEGDLWL-----AAS 633  
Db 630 -----SNNIGFANTGS-----NNIGLGTGCGIG--FGLNSGSGNIGLFGNSGNI 675  
Qy 634 GLYHSTNG-----GSSWSAITGV--SSAVNVGFGKSA-----PGSSYP 669  
Db 676 GFFNSGNGNVGIGNTGTANFGLGNTGTNTGFFNSGDVNTGIGTGSFNTGSFNPDSNT 735  
Qy 670 AVF-----VVGTIGGV--TCAYRSDDC--GTTWVLINDDQHQYGNWQCAITGDHANL 717  
Db 736 GDFNPGSYNTGLGNTGVDVTGAFISGYSNGFLW-----SGNY-QGLIGLHAAL 783

RESULT 6  
US-11-052-554A-142  
; Sequence 142, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 142  
; LENGTH: 3157  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-142

Query Match 3.9%; Score 156; DB 7; Length 3157;  
Best Local Similarity 20.3%; Pred. No. 0.012;  
Matches 174; Conservative 96; Mismatches 295; Indels 294; Gaps 42;

Qy 13 IGGG----GFDVGIVFNEGAPGILYVRTDYGMYRWDAAAGRWIPLLDWVGWNNWYNGV 68  
Db 811 VGGNVGNGIGDANFGLNAGLAGACVGNIGLGNAGSN-----VGFNGWGVNI 863  
Qy 69 VSIADPINTNKVAAVGMVMTNSWDPNDGAILRSSDQGTWQITPLPKLGGNMPGRG-M 127  
Db 864 -----GFGNTGNNLIGLGTGDNQGT-----IGLNSGAGNI 895  
Qy 128 GERLAVDPNNDILYFGAPSGK-GLWRS-----TDSGATWSQ-MTNPPDVGTVIANP-- 177  
Db 896 G---LFNSGTGNVGLFNSGTGNTGFLNSGSGFNTGIGNGTGTGLFNAGNFTGVANPGS 952  
Qy 178 -----TDTTGYSQSDIQGVVWVAFDKSSSILGQASK----- 207  
Db 953 YNTGSFNVGDTNTGFGNPGSINTGW--FNTGANTGVANSNVDTGALMSGNFGNIGLR 1010  
Qy 208 -----TIFGVADPNPNPFWSRDGGATWQAVPCAPTGFPHKGVPDPVNHVLIATSN 260  
Db 1011 GNPEGLFGLNVGIIPEPPIHWTSTGGI-----GPI-IIPDTTILPPI-HGLTGOAN 1061

QY 261 TGGPYDSSGDVWKFSTSGTWRISPVSTDTANDYFGYSLTIDRQHPNTIMVATQIS 320  
Db 1062 YG-----FAVDP-----IIPAI-----HIDFG----- 1080  
QY 321 WWPDTIIPRSTGGATWTRIMWTSYPNRSRYVLIDISAE-----PWLTFGQPNPPVSP 376  
Db 1081 -----ADAGPT-----APATLLSALGITGQPRFGPITVSNVQLNPFNVNL 1122  
QY 377 KLGWDEMA---IDPFNSDRMLYG---TGATLYATNDLTKWDSGGQIHIAPMVKGLEET 430  
Db 1123 KLOFLHDAFFNEFPDPTISVQIOVAIPLTSATL-----GGLALPLQOTIDAIELP 1172  
QY 431 AVNDLISPPSGAPLIS-ALGDLGGTHADVTAVPSTIFTSPVFTTGTSDVYAEALNPSIIV 489  
Db 1173 AIFSQSIPIDIPIDIPASTINGISMSEVPIDVSDIPAVTIITGTIRIDPLNFDVLS 1232  
QY 490 RAGSPD-----PSSQPNDRHVAFTSDGKNWFGQSEPGGVTTGGTVA 531  
Db 1233 SAGPINISIIIPALPGFNSTELPSSG-----FFNTGGG-----GSGGI 1272  
QY 532 ASADGSRFVWAPGPGQPVVYAV-GFGNSWAASQGVANAIQIRSDRVNPKT----- 581  
Db 1273 ANF-GAGVGLLNQOASSPMVGTLSGLNAGSLASGLVNSGVDISGMFNVSTLGSAPAVIS 1331  
QY 582 -FYALSN-----GTFYRSTDGGVTQPVAAGLPS--SGAVGVMFHAPV----- 621  
Db 1332 GFGNLGNHVSIGVSDGLLAMLTSQG-----SGSGQPSIIDAAIAELRLNPLNVLNGLV 1387  
QY 622 -----GKEGDLMAASSGLYHSTNGSSWSAITGVSSA--VNVFGKSAFG-----S 666  
Db 1388 GSYNLGFANVDGNLGNAGNLGNLGNLGNLGNLGNLGNLGNLGNLGNLGNLGNLGN 1447  
QY 667 SYPAVVVGTIGGVTGAYRSDCGTTWVLINDDQHOYGNWQAITGDHANLRVYVIGTNG 726  
Db 1448 GLGALPGIGNLG--LGNAGSNVVG-----FCNMGL-----GNIGFGNTGNN 1487  
QY 727 RGI-VYGD-----IGGAPSG 740  
Db 1488 LGIGLTGDNQTFGGLNSG 1506

## RESULT 7

US-11-087-099-7521  
; Sequence 7521, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 7521  
; LENGTH: 7465  
; TYPE: PRT  
; ORGANISM: Magnetococcus sp. MC-1  
US-11-087-099-7521

Query Match 3.8%; Score 155; DB 7; Length 7465;  
Best Local Similarity 19.7%; Pred. No. 0.041;  
Matches 153; Conservative 69; Mismatches 287; Indels 268; Gaps 31;

QY 76 INTNKNV-----AAGVMTNWDPNNDGAILRSSDGA---TWQITPLPKLGNWPG 124  
Db 6509 VGSNGVMSYQLDNSKAVVQGLTQGOQVSESFTVHTSDGGSQSLTNIT----- 6556  
QY 125 RGMGERLAVDNNILYFGAPSGKGLWRSTDSGATWSQMTNFPDVGTYIANPTDTTC-- 182  
Db 6557 -GRQDSAVVSGTGVSTEDSTLSSSGMLSVSDADSGEAAVYASTQTGSGYFSLDSSGAW 6615  
QY 183 -YQ-----SDIQGVVWVAFDKSSSLG-QASKTIFVGVADPN-NPVFWSRDGGATWQAVP 234

Db 6616 SYQLNNSASQVQALI-----MGQVSBESFTVATVDGTASSVVVTVVGAQDAALQLG 6665  
QY 235 GAPTFGIPHKGVDFPNVHV-----LYTATNTG--GPYDSSGDVWKFSTVSTGTW 282  
Db 6666 GYQSGVVEDGQLSTAGQLTISDADAGSLYQVASTDGMYGHPAMDSSGAWSY-VLNNE 6724  
QY 283 TRISVPVSTDTANDYFGYS-----GLTIDRQHPNTIM---VATQISWWDITIFRS 330  
Db 6725 AAIQMIAGTQVLESFTVSSSDGTQAVSVTITGSDNAVITGEEAAGVSETNDDSSMA 6784  
QY 331 TD-----GGATWTRIMWTSYPNRSRYVLIDISAEPLWTFGVPQPNPPVPS 375  
Db 6785 TCKLINSIDLSQOAHFVGPGQAGVCAFTLOGDGNWTVULD-----NHGV----- 6829  
QY 376 PKLGMWDEMAIDPFNSDRMLYGTGATLYATNDLTKWDSGGQIHIAPMVKGLEE- 429  
Db 6830 -----AMA-----GLTGDVVTFTVRSILDNTAETVITTINGLDEPPTAE 6872  
QY 430 -----TAVNDLIS-----PPSGAPLISALGDL----- 451  
Db 6873 IEARLABEORLQBELQASNDLLOGAIGAEGGNAGGTAPPNPNGNPADGAPDGLGAGGA 6932  
QY 452 -----GGFTHADVTAVPSTIFTSPVFTTGTSDVYAEALNPSIIVRAGSDPSSQPNDRHV 505  
Db 6933 AAGGATGGTGGTTA-----GTGGGGLG-----GGFG-SCAPNT--- 6967  
QY 506 APTDGGKNWFOGSEPGGVTTGCTVAASADGSRFVWAPGDPQPVVYAVGFGNSWAASQ 565  
Db 6968 --SAGLTTGGGTATGTTTGT----- 6998  
QY 566 VPANAQIRSDRVNPKTFYALSNGTTFYRST--DGVTFQPVAAGLPSSGAVGVMFHAPVGK 623  
Db 6999 GTA-----TGCTATGTTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 7034  
QY 624 EGDLMALASSGLYHSTNGSSWSAITGVSSAVNVNFGKSAAGSSYPVFTVGTIGVTGA 683  
Db 7035 AGGTGGAAGGAATGEAGGTATGCTGTTGGAATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 7089  
QY 684 YRSDDCGTTWVLINDDQHOYGNWQAITGDHANLRVYVIGTNGRGIYVDIGGAPSG 740  
Db 7090 AATGEAGGT-----ATGCTGTTGTTGTTA-----TGTTGGAAGTGGAAAG 7129

## RESULT 8

US-11-087-099-12161  
; Sequence 12161, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 12161  
; LENGTH: 807  
; TYPE: PRT  
; ORGANISM: Cytophaga hutchinsonii  
US-11-087-099-12161

Query Match 3.8%; Score 153.5; DB 7; Length 807;  
Best Local Similarity 17.9%; Pred. No. 0.0034;  
Matches 160; Conservative 108; Mismatches 223; Indels 403; Gaps 52;

QY 8 WSNVAIGGGGFGVDGIVFNEGAPGILYVRTIDIGMYRWDAANGRWIPLLDWVGWNNW--- 63  
Db 31 WKWA-----KGSDDHTTAVRQD-CTLNSW-----GQMYGQLG 62  
QY 64 ---GNGVVSAAADPINT-NKVMAAVGMYTNWDPNNDGAILRSSDQATW----- 109  
Db 63 DGSYSSIIIPVQVGLTWTWVKVFA---KYDNSF-----AIMK---DGTLMWGMNSLIGTL 111  
QY 110 -----QITPLPKLGNWPGRCMGERLAVDPNNNDILYFGAPSGKGLWRSTDSGATWSOM 164



Db 112 GDTQVRKTPVKV-----TATD-----WQSV----- 135  
 Qy 165 TNFPDVTYIANPTDTTGYQSDIOGVVWVAFKSSSLGQASKTIFVGVADPNPWFMSR 224  
 Db 136 TNF-----NFTAIAKKD--GSLWMDNYNGQLNGTKV-----KNLPSKV 175  
 Qy 225 DGGATQAVP--GAPT-----GPIPHKGVFDPV----- 250  
 Db 176 GSGKNWKSVSIGSSNLAIKTDGTLWGMNMYNQIGYPTVDVLSPQTQIGSETDKWSVS 235  
 Qy 251 ---NHVLYIATSN---GGPYDGSSGD-----VWKFVS 277  
 Db 236 CGDHHVAIRTGLSLWTSKQIGNGEINSVQTEPVQVSGSDTDMGTCHAIQDWSIAI 295  
 Qy 278 -TSGT-WTRISFPSTDTANDFYGSGLTIDRQHPNTIMVATOI-----SW---WP---DT 325  
 Db 296 KTGSLWMTWGS-----YGYNSSYIRKN-----IPTQGRDTSWMIYPSQOS 340  
 Qy 326 IIPRSTGGATWTRIDWTS-----YPNRSRLRYLD-----ISABPW-----L 363  
 Db 341 VAFK--TDGS-----LWTGWSNARQLGLGTYCNTNIPYVWSCPEVLTSQWQAASIVQST 394  
 Qy 364 TFGVQPNPPVPSPKLGMWDEAMADFPNSD-----RMLYGTGATLYATNDLTK 411  
 Db 395 SFGURTDSTL--WKWGWGNE--NSLDIYARNPDGPFARIGTYKRAYASGGALLAIKDDGTL 451  
 Qy 412 WDSGGQIHIAPMVKLEETAVNDLISPPSGAPLISALDGLGTFHADVTAVPSTIFTSPV 471  
 Db 452 WQSGR-----NVPVL----- 462  
 Qy 472 FTTGTSVDYAEINPSIIIVRAGSFPDPNDRHVAFSTDGG-----KNMFQSGSEPG-- 523  
 Db 463 --LSSSKDWKAAVAS-----SSRGVAIKTDGSLWMDIDNYNNTEPSGGV 506  
 Qy 524 -----VTGTGVAASA--DGSRFWAPGDPQPVVYAVGFNSWAAQGV 566  
 Db 507 IPLLKEINPGTQWQSISASTTTAAAIARDGSLWIW-----GSALYGAMGTG---IAVAGS 558  
 Qy 567 PANAOIRSDRNPFTYALSNGTFYRSTGDTGVTQPVAAAGLPSSGAVGVWFMHVPKGEK 626  
 Db 559 PTLIQMND-----TGQFVSVESTIWAIKT-----DGT 588  
 Qy 627 LWLAASGLYHSTNGSSWSAITGVSAVNVVFGF---KSPAGSSYP--AVFVGTIGGVT 681  
 Db 589 LWACGQNNYQGLNGNT-----TDIYTLTQIGTATDMKTVVAGPYHTLAIKTDGSING-W 642  
 Qy 682 GAYRSDCGTTWVLINDDQYQWNGQAITGDHANLRVYGTNGRGIVY--GD 733  
 Db 643 GSYTYNKLGLGYDLYGNLLEQ--NWKHIATGPYNS-----FAIHGDCGTLVTVGD 689

RESULT 9

US-11-087-099-11214  
 ; Sequence 11214, Application US/11087099  
 ; Publication No. US20060041961A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S. et al.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53450)B EP  
 ; CURRENT APPLICATION NUMBER: US/11/087,099  
 ; CURRENT FILING DATE: 2005-03-22  
 ; NUMBER OF SEQ ID NOS: 12464  
 ; SEQ ID NO 11214  
 ; LENGTH: 894  
 ; TYPE: PRT  
 ; ORGANISM: Cytophaga hutchinsonii  
 US-11-087-099-11214  
 Query Match 3.7%; Score 148.5; DB 7; Length 894;  
 Best Local Similarity 20.0%; Pred. No. 0.0086;  
 Matches 145; Conservative 87; Mismatches 238; Indels 255; Gaps 39;

Qy 7 TWSNVAI-----GGGFVDGIVF-----NEGAPGI-----LYVRTDIGMYRDAAN 48  
 Db 203 TWTGIAACFSSGIGIKANGTLTWGPNNGGELGIGNDEKHTPVQVQADQNMUYAVAGSN 262  
 Qy 49 GRWIPLLDVGWNNWNGYVWSI-----AADPINTKVAAVGMVTN----- 90  
 Db 263 HIVALKVDGSLW--GWGLNEAGEIGNGTNVQODSPVRVGTIDNDWTTTLAGSNHTLAIKSNG 321  
 Qy 91 ---SWDPNDGAILRSSDQGAQTWQITP--LPFKLGGNMP-----GRGMRGLAVDPNND-- 138  
 Db 322 TLWAWGHN---VRGNLGNATLDPATLPVQVGTADWIRVSAGLDHSAVALKANGTLWA 377  
 Qy 139 ---NI---LYGAPSGKGLWRTSDSGATWSQMTNPDVGTIYIANPTDTTGYQSDIOGVVW 192  
 Db 378 WGVNVRGOLGYGSGTDKHLVPQIGTAHTWVAI-----NAGTY-----HTIGVKAD--GSLW 426  
 Qy 193 VAFKSSSSILGOASKTIFVGVADPNPWFMSRDCGATWQAVP---GAPTGTIPHKGVFD 248  
 Db 427 TWGNTNAQLG-----DGGTADQPVPHSIRTADDDWIS----- 459  
 Qy 249 PVNHVLIATSNTEGPGYDGSSGDVWKFVS---TSGTWTRISFPVSTDTANDFYGSGLT 304  
 Db 460 -----IAASNAFSMGLKANGTIWGLNPFETDGOYKNSSPV---QTGSD----- 501  
 Qy 305 IDROHPNTIMVATOISWMPDTHIIFRSTGGATWTRIDWTSYPNRSRLRYVLDISAEPWLT 364  
 Db 502 -----QNMKSIATGSNY---ILALKADG---TLWAWGD-----N 529  
 Qy 365 FGVPQNPVPSPKL-----GHWDEAMALDPNSDRMLYGTGATLYATNDLTKW--D 413  
 Db 530 FSGQLGCGTEQPRMLPKOIGTATTWVSIAAA---SGVQSFG---IQADGSLMSWGHN 580  
 Qy 414 SGGQIHIAPMVKLEETAV---NDLISPPSGAPL---ISALDGLGTFHADVTAVPSTIF 467  
 Db 581 YDGLGIGTNLKTLVPTFVGTGTWISIKTVGHTLGIKSDGTLWGW----- 627  
 Qy 468 TSPVFTTGTSDYVYAEINPSIIIVRAGSFPDPNDRHVAFSTDGGKNWFGSSEPGVTTG 527  
 Db 628 -----GTN-----RQQLGDGTLVN--KYVPVQLGTDTRDWIN-----IAG 660  
 Qy 528 GT---VAASADGSRFVWAPGDPGQPVVYAVGFNSWAAASQGVPANAOIRSDRNPFTFYA 584  
 Db 661 GTIHNICLKSGTLLWAGRNNGYQ---VGTGNLITQDRMVQISAE---QNMIDVYA 710  
 Qy 585 LSNCTFYRSTGGVTFQPVAAAGLPSSGAVGVWFMHVPKGEK-----LWLAASS 633  
 Db 711 SLDOQFAIRADGSLW---ACGLNSSGQLG-----DGTSDVRAAMVAVGTNQKRWISLAA 760  
 Qy 634 GLYHS 638  
 Db 761 GNYHT 765

RESULT 10

US-10-517-939-330  
 ; Sequence 330, Application US/10517939  
 ; Publication No. US20060003433A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steer, Brian  
 ; APPLICANT: Callen, Walter  
 ; APPLICANT: Healey, Shaun  
 ; APPLICANT: Hazlewood, Geoff  
 ; APPLICANT: Wu, Di  
 ; APPLICANT: Blum, David  
 ; APPLICANT: Esteghalalian, Alireza  
 ; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
 ; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
 ; FILE REFERENCE: 56442007901  
 ; CURRENT APPLICATION NUMBER: US/10/517,939  
 ; CURRENT FILING DATE: 2004-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/US03/19153  
 ; PRIOR FILING DATE: 2003-06-16  
 ; PRIOR APPLICATION NUMBER: 60/389,299

; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq For Windows Version 4.0  
; SEQ ID NO 330  
; LENGTH: 755  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(35)  
US-10-517-939-330

Query Match 3.7%; Score 147.5; DB 6; Length 755;  
Best Local Similarity 20.4%; Pred. No. 0.0082;  
Matches 159; Conservative 74; Mismatches 254; Indels 291; Gaps 41;  
QY 44 WDAANGRIPIPLDVGWNNWNGVGVSTAAADPINTKVAAGVMTNS---WDPNDGAIL 100  
DB 123 WHSQNPGLTN-----GNWSRSQLINIMNDHIDT-----VAGRYAGEVLVWD---VYN 167  
QY 101 RSSDQATWOTPLPFLKGGNMPGRGMRERLAVDP-----NNDNLYFGAPS----- 147  
DB 168 QAFNEDGTYRSTIYNGIGOEYIDIAFTRARAADPHAKLIYNDYINIGLNSKSNVYNA 227  
QY 148 -----GKG-----LWRSTDSGATW-SQNTNFPDVG--TYIA-----NPTDTT 181  
DB 228 ADMVRGVPIDGVGFQMHLERGVSGSSLASNMORFADLGLVYITELDVRIPQNT--- 284  
QY 182 GYQSDIQGVWVAFDKSSSLGQ-ASKTIFVGVADPNPNPVFWSRDRGGATWQAVP----- 234  
DB 285 --QODLOQAAYQVQVNRCLAQAPACKALQV-----WGIPDKYSW--VPDVFPGT 330  
QY 235 GAPTGFIFH---KGVPDPVNHVLYIA-----TSNTGPPVDGSSGDVWKFVSIGTWTRISP 287  
DB 331 GAPLLFNNDYBAKPAYAVQAEQMAANPQNTPGTPAHTPS-----ATSTSAATATPP 384  
QY 288 VPSDTANDYEGYSGLTIDRQHPNTIMVATQISWPDITIIIPRSTGGATWTRINDWTSYP 347  
DB 385 ATATATATPSSGGGVCAVD----- 403  
QY 348 NRSRLYVLDIAEAPWLTGEGVQPN-----PPVSPKLGWDEAMADPNFSDRMLYGT 399  
DB 404 -----YVI---ANQWGN-GFOANVTIINHSAPVNGYTLAWTHAPQOIVTSQWNTIAQS 454  
QY 400 GATLYATNDLTKWD-----SGQIHL-----IAPMVKGLBEETAVNDLISPPSGA 442  
DB 455 GSASASNPAGYNGVIGANGGKISFGQSLAGGSAAVAPTYFALNGAACNGAVLPPTAT 514  
QY 443 PLISALGDLGGFTHADVT-----AVPSTIFTSVPFTTGTSDVYAEINPSIIVRAGSPDPS 497  
DB 515 -----FTSPPTATMCPQATPELLVQPVTSPTT-----QLSQTLLVRLNGSB-- 556  
QY 498 SOPNDRHVAFTDGGKWFQSGEGVTTGGTVAASADGSRFVWAPGDPQPVVYAVGFG 557  
DB 557 -----WRAAGPAGVVT-----VTAP-DPD-----GYF 578  
QY 558 NSWAASQGVPAQAIRSDRV-----NPKTP--YALNSGTIFYRSTDDGGVTFQPVAAGLPS 609  
DB 579 RLTIPLAANTSNAILLVEGRVVRVITHSNGCTYGGYTLSTRTVTVIQVASSPVLTPATPSPT 638  
QY 610 -----SGAVGVNF-----HAPVQKE--G 625  
DB 639 ATATPTVTATPSGACTVAYAITNDWGSFGFTANVTLTNTGSSALMGWTLAYAFPGNQDIS 698  
QY 626 DLW--LAASSGLYHS-TNGGSSWSAIGTGVSSAVNVGFKGSAPG--SSYPAVFVVGITGG 679  
DB 699 NAWNGTAVQSGSSSVTVNAG--WNGSLPPNVASAFQOASYSNGNSVSPASF---TLNG 751

RESULT 11

US-11-052-554A-6

; Sequence 6, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 949  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-6

Query Match 3.6%; Score 146; DB 7; Length 949;  
Best Local Similarity 20.0%; Pred. No. 0.014;  
Matches 147; Conservative 77; Mismatches 252; Indels 258; Gaps 37;  
QY 136 NNDNLYFGAPSG---KGLWRSTDS---GATWSQMTNFPDVGTYIANPTDIT--GYQSD 186  
DB 70 NHDNQIVCTANGMTITSTGLGLPDSSEENTGGQIQN-----GGIAGNTVTTNGRQVV 123  
QY 187 IQGVWVWAFDKSSSLGQASKTIFVGVADPNPNPVFWSRDRGA-----T 229  
DB 124 LEG-----GTASDTVI-----RDGGQSLNGLAVNTTLNRRGEQ 157  
QY 230 WQAVPCAPTGFPHKGVFPDPVNHVLYIATNTGPPVDGSSGDVWKFVSIGTWTRISPVP 289  
DB 158 WYHEGVAFTGTINRDGYQSV-----KSGG-----LATGTIINTGAEG 195  
QY 290 STDITANDYFGY-----SGLTIDRQHPNTIMVATQISWPDITIF-----RSTDGGAWTWR 339  
DB 196 GPDSDNSYTGQKVGQVTAESTTINKNGRQIILFSGLAR---DTLIIYAGDQSVHGRALNTT 252  
QY 340 IWDTSYNRSLRYVLDIAEAPWLTGEGVQPNFPVSPKLGWM----- 381  
DB 253 LGGGYQVHRD-----GLALNTVINEG--GWQVVKAGGAAGNTTINQNGE 295  
QY 382 -----DEMAIDPFPNSDRMLYGTGATLYATNDLTKW-----DSGQIHLAPM 423  
DB 296 LRHVAGGEATAVTQNTGGALVTSTAATVIGTNRNLGNFTVENGGKADGVVLESGRGLDV--- 352  
QY 424 VKGLEETA VNDLISPPSGAPLISALGDLGGFTHADVTAVPSTIFTSPVFTTGT-----S 477  
DB 353 ---LSSHSAQNTLVDGDTGLAVSAGKA---TSVITISGALLIADSGATVEGNTASQKFS 406  
QY 478 VD--YAEALNPSIIVRAGSPDPSQPNDRHVAFTDGGKWNFGQSGEGVTTG--GTVAAS 533  
DB 407 IDGTSQASGLLENGSGSFVNA-----GGQAGNTTVHGRGTTLA 447  
QY 534 ADGSRFVWAPGDPQPVVYAVGFGNSWAASQGVPAANA-QIRSD-RVNPKTFFYALNSGTFFY 591  
DB 448 AGGSLSGRQTQSKGASMLV-----NGDVVSTGDIVNAGEIRFDNQTPNA--ALLSRV-- 498  
QY 592 RSTDGCVTQPV-AAGLPSSGAVGVWFAVPGKEGLWLAAASGLYHSTHSGSSWSAITG 650  
DB 499 AKSNSPVTFFKLTITNTLTQGGGTINMRVLDGNSASDQLVINGG--QAT--GKTLTAFTN 554  
QY 651 VSSAVNVGFKGSAPSSYPVAVFGVGTIGVVT-----GAYR---SDCCGTW 693  
DB 555 VGNS-NLGVATTGQG-----IRVVDQNGATTGEGAFALSRPQAGAFNTLNRDSDWD 608  
QY 694 -----VLINDDDHQYGNWGOAITGDHANLRVYIGTNG 726  
DB 609 YLRSENAYRAEVPYLTSMLTQAMDYDRILAGRSRSHQTG-----VNGENNSVR---LSIQG 660

727 RGLVYGDIGGAPSG 740  
:  
:  
:  
661 GHLGHDNNGGIARG 674

RESULT 12  
US-11-052-554A-133  
; Sequence 133, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 133  
; LENGTH: 3300  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-133

Query Match 3.6%; Score 145.5; DB 7; Length 3300;  
Best Local Similarity 21.5%; Pred. No. 0.069;  
Matches 183; Conservative 87; Mismatches 299; Indels 281; Gaps 48;

QY 10 NVAIGGGGVGDGIVFNEG-----APGILYVRTDIG---GMRYMDAAN-GRWIPLLDWVGW 60  
DB 2467 NANIGCANICD---FNVGIA NTGELTA AVNNIGNTGN YNICVGNTGNYN-----LGF 2518  
QY 61 NNWGYNGV-VSIIAD-----PINTNKVAAVGYMTNSWDPN DGAILRSSDQGATWOITP 113  
DB 2519 GNTGNNNIGLSGDNQIGFGPLNAG--IANMGL----- 2550  
QY 114 LPEKLGDN--MFGRGWERLAVDPNDNLIYFCAPSGK-GLWRSTDSGATWSOM----- 164  
DB 2551 --FNLGDNFMGMANAGNFNOGIANTGNNNI GLFNTGNNNVGIMLTGDLGSFSSLNSGAG 2608  
QY 165 -TNFPDVGVTYI ANP-----TDTTGYOSDI QGVVVWA----- 194  
DB 2609 NTGCFNFSGT--ANTGLFNSGTGNTGLFNSGTGNGVIGNMGTGCGVGLSGDSQVIGGTTN 2666  
QY 195 -----FDXSSSI LGQA-SKITFVGADVPPNPWFVSRDG-GATVOAVPGAPTGFPHK 244  
DB 2667 SGSFNIGLFNSGTGNGVIGNSGTGNGVIGNTGT---GNTGIGNSGNYNTGLLNAGLVNT 2722  
QY 245 GVFPDPVNH---VLXIATSNTG---GPYDSSGDVWKFSVTSGTWTRISPV PSTDTANDY 297  
DB 2723 GIANGPNHNTGLFNIGTFNTGIANP GHYN-----TGSYNTG SYNTGMANAGDY 2770  
QY 298 FGYSGLTIDRQHPTIMVATQISWWPD TIIFRSTDGGATWTR-----IWDTSVPYRSRL 351  
DB 2771 -----GTCAFITGSNNWGLLRADROQLLAANYTTITERPA 2806  
QY 352 RVV-LDISAEPMWITFGVQP NPVPSPKLGMWDEAMAID-PFNSDRMLYGTG-ATLYATND 408  
DB 2807 AFLNVLDIPVNIPI T-GDITNVSI PAITFRIDASGSVDIGILSGTVLAPVGPITLHGDA 2865  
QY 409 LYKWDSGGQIH IAPMKVLEETA VNDLI SPSPCAPLISALGD LGGFTHADVTPSIFT 468  
DB 2866 SAPLDTPIEIDFGP-----SPA INLIGKPDGSTVINIVGGAG-----A 2904  
QY 469 SPVF TTGTSDVYAELNPS II VRAGS DPSSOPNDRHVAFST DGGKNWFQSGEPGVITGG 528  
DB 2905 GPI-----SIP IIDLRA-----PGFFVATTGP-----SSGFLNWGAGSACLLNFGN 2947  
QY 529 T-----VAASADGRSFV-----WAPGD PGQPVVYAVVFGNS-----WAASGQVPANA 570

|  |      |  |      |
|--|------|--|------|
| Db   | 2948 | NSGLYNFATSSMGNSGFONYGSLQSGWA-----NLGNSISGIYNTGLGAPANV          | 2999 |
| Qy   | 571  | QIRSDRWPKTFYA--LSNGTFYRSTDDGGVTFQPVAAGLP-----SSGAVG            | 614  |
| Db   | 2996 | -----SGLLINTGNLWGLQNG-----PTE--TTFSVGLANLGFNWLGSANIGNYNLGSANIG | 3046 |
| Qy   | 615  | VMFHAVPGKEGDLWL-AAASGLYH--STNGGSSWSAITGVSSAVNVGFGKSAPEGSSYP    | 671  |
| Db   | 3047 | V-YNLGSANIGDFNLGSANIGDFNLGSANIGSS-----NIGFGNVGFG-----LT        | 3090 |
| Qy   | 672  | FVVGTTG-GVWGAYRSDDCGTTWVLINDHOHQVGNMGQAITGTHANLRRVYICTNGRGIV   | 730  |
| Db   | 3091 | AALICNIGFGTNGNIGCNIGTGTGTCNIGCNIGICGLTGD-----TMTGFGGWSNG       | 3144 |
| Qy   | 731  | YGDIGGAPSG 740   |      |
| Db   | 3145 | TGNIGLFNSG 3154  |      |
| RESULT 13  |      |  |      |
| US-11-052-554A-68  |      |  |      |
| ; Sequence 68, Application US/11052554A  |      |  |      |
| ; Publication No. US20050288866A1  |      |  |      |
| ; GENERAL INFORMATION:   |      |  |      |
| ; APPLICANT: Sachdeva, et al.  |      |  |      |
| ; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN |      |  |      |
| ; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL                        |      |  |      |
| ; FILE REFERENCE: 30853/40359A   |      |  |      |
| ; CURRENT APPLICATION NUMBER: US/11/052,554A                                   |      |  |      |
| ; CURRENT FILING DATE: 2005-02-07  |      |  |      |
| ; PRIOR APPLICATION NUMBER: US 60/589,227                                      |      |  |      |
| ; PRIOR FILING DATE: 2004-07-20  |      |  |      |
| ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004                                    |      |  |      |
| ; PRIOR FILING DATE: 2004-02-06  |      |  |      |
| ; NUMBER OF SEQ ID NOS: 763  |      |  |      |
| ; SOFTWARE: Seqin version 3.3  |      |  |      |
| ; SEQ ID NO 68   |      |  |      |
| ; LENGTH: 760  |      |  |      |
| ; TYPE: PRN  |      |  |      |
| ; ORGANISM: Escherichia coli 0157:H7   |      |  |      |
| US-11-052-554A-68  |      |  |      |
| Query Match 3.6%; Score 145; DB 7; Length 760;                                 |      |  |      |
| Best Local Similarity 20.6%; Pred. No. 0.012;                                  |      |  |      |
| Matches 137; Conservative 77; Mismatches 194; Indels 258; Gaps 40;             |      |  |      |
| Qy   | 156  | DSGAT--WSQMTNFPDPVGYIA--NPTDITGYQSDIQGVVWVAFDKSSSLGOASKTIFV    | 211  |
| Db   | 85   | DQCATNLTLDAKNVPGVGAFFAGENGNSITGDAIYMRG-----ADTSNSIYI           | 131  |
| Qy   | 212  | -GVADPNPVPFRDCGATWQ--AVPG-----APTGFPHKGVPDPVNHVLIATS           | 259  |
| Db   | 132  | DGIRDIGSV---SRDTFTEQVEVIKPGSGTDYGRSAPTG-----SINMISKPRN         | 179  |
| Qy   | 260  | NTGCGPDGSGDVKWKFSTVSGTWTRISPVSTDTA-----NDYFGY                  | 300  |
| Db   | 180  | DSGIDASASIGSAW---PRRGT-LDVNQVIGDTTAVRLNVGMGEKTHDAGDKVKNERYG    | 235  |
| Qy   | 301  | S-----GL-TIDQHPNTIMVATQISWPPDTTIFRSTDGATWTRIDWMTSPNRSRYV       | 354  |
| Db   | 236  | APSIAPFGLGTANRLYNLHV-TQ-----HNTPDGGIP-----                     | 267  |
| Qy   | 355  | LDIASFPWITFGVQPNPPVPSPKLGHMDDEMAIDPNSDRMLYGTG-----ATLY         | 404  |
| Db   | 268  | -----TIGL-PCYSAPSAGTATLNHSGKVDYTN-----FYGSDSDYDDSTDTATMR       | 313  |
| Qy   | 405  | ATNDL-----TKWDSGGQIHIAPMWKGLEETAVNDLISPSPGAPLISALDGLGGFT       | 455  |
| Db   | 314  | FEHDINDNTTIRNTRWSRVKQDYLMTAIG-----CASNITQPTS-----DVSNT         | 360  |
| Qy   | 456  | HA-----DV-----TAVPSTIFTSPV-----FTTGTSTVDYAEALNPSIIVRAG         | 492  |

Db 361 WSRANTKDYNSKILTNQTNLTSTFYTASIGHDYSTGVETRETQTNVG-VNPVTLPAVN 419  
Qy 493 SFDSSOPDRHVAFTDGGKQNFQGSBPGVTTGGTVAASADGSRFVWAPDGPQPVY 552  
Db 420 IYHPDS-----SIHPGGLTRNG--ANANGTDTFA-----IY 449  
Qy 553 AVFGNSWAASQGVPANQAIRSDRVNPKTFYALNSCTFYRSTDDGVTQFQVAAAGLPSSGA 612  
Db 450 AF---DTLOITRDPFELNGGRLNDNYHTEYDSATACG---GSGRGAIT---CPAGVAKGSP 500  
Qy 613 VGVNFHVPKGGDL--WLAASSGLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSVPA 670  
Db 501 VTTVDTA---XSGNLVNWKAGA--LYHLTENVY-----INAVSQPPGGN-- 544  
Qy 671 VFVVGTTGGVTGAYRSD-----DCGTTWVLI-----NDD- 699  
Db 545 -FALAQSGSGNSANRTDFKPKQANTSEIGTKWQVLDRKLLLTALFRDIEVEVEQNDG 603  
Qy 700 -HQYQ 704  
Db 604 TYSQYG 609

RESULT 14  
US-11-052-554A-179  
; Sequence 179, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 179  
; LENGTH: 955  
; TYPE: PRT  
; ORGANISM: Shigella flexneri 2a str. 2457T  
US-11-052-554A-179

Query Match 3.5%; Score 140.5; DB 7; Length 955;  
Best Local Similarity 20.1%; Pred. No. 0.033;  
Matches 162; Conservative 98; Mismatches 309; Indels 237; Gaps 38;  
Qy 77 NTNKVMAVGMVYNSWDPNDGAILR-----SSDQATWQITPLPKLGG- 120  
Db 178 NSNGIGTINDYGTSDVLSGSKIKTQGTGVYIGGLNGNANGAARPTATDLTIDVOG 237  
Qy 121 NMPGRGGERLAVDPNNDNIIYFCAPSKGLWR--STDGATWQMTNFPDVGVIANPT 178  
Db 238 SAMGINVQKNSVDLTGNTSIKTNDAHGLWFGQVSANALTVDTGAAANGVEVRGT 297  
Qy 179 DTTGYQSDIQGVVWVAFDKSSSLGQASKTI-FYGVADPNPNPFWSRDGGATWQAVFGAP 237  
Db 298 TTIGADSHISSAQGGGLVTSSS-----DATINFSGTAAQRNSIF-----SGGSYGASAQTA 348  
Qy 238 TGFPHKGVDPVNHVLIAT---SNTGPGVDSGSDVWKFVSVTGWTTRISVPSTDTA 294  
Db 349 TAVINQMONTITVDNRGSLALGLWALSGRITGDS-----LAITCAAGAR-GIYAMTNSQ 402  
Qy 295 NDYFGYSLRTDROHPNTIIVATQISWMPDPIIFRSTDDGATWTRINDWTSYPRSLRYV 354  
Db 403 IDL---TSDLVDMSTPDQMATQ-----HDDGYAASRI-----NASGRL 441  
Qy 355 LD---ISAEPLWTFGVOPNPVPSPKL-----GWMDEAM----- 385  
Db 442 INGSVLKGGILNDLHDPGSVWTCSSLSNDVNGKLDVAMNNSVWNTVNSNLDTLALSH 501

Qy 386 -AIDPFNSDRMLYGTGATLYATN-----DL--TKWDSGGQI 418  
Db 502 STVD-FASHGSGTAGTFTLNLENLSGNSTFIMRADVVGEGNGVNRGDLNLTSGSAGNH 560  
Qy 419 HIAPVVKGLETAUND--LISPPSGAPLISALG--DLGGETH----- 456  
Db 561 VLAINQSEATTTGNEVLTVVKTDTGAAASFSSQVELGGYLDVRKNTNNWELYASGTV 620  
Qy 457 -----ADTAVPSTIFTSPV-----FTTGTSDVYAEALNPSII 488  
Db 621 PEPTNPPTTAPAPQPIVNPDPTEPAPTPTTADAGNYLVNGLLYNVE-NRTL 679  
Qy 489 VRAGSFDSSPNDRHVAFSTDGK--NWFQSEPG-----GVTTGG----- 528  
Db 680 QRMG--DLRNQSKGNIWLRSYGGSLDSFASGKLSGDMGYSGIQGGDKRLSDVMPLYV 737  
Qy 529 -----TVAASDGRFVWAPDGPQVYVAVYVFGNSWAASQGVPANQAIRSDRVNPKTFY 593  
Db 738 GLYIDSTHASPDYS-----GGDGTARSDYMGMYASYNAQNGFYSDLVIKASR-QKNSFH 790  
Qy 584 ALSNGTFYRSTDDGVTQFVAAAGLPSSGAVGVNFHAPV-----GKGGDLW 628  
Db 791 VLD-----SONGVNANGTANGMSISLEAGORFNLSPTCGYFIEPQTLTYSHQENMA 844  
Qy 629 LAASSGLYHSTNGSSWSAITGVSSAVNVGFGKSAPGSSYPVAVVGTIGGTGAYRSD 688  
Db 845 MKASNGLIHLN--HYESLLGRASMI-LGYDITA-GNSQLNVYV-----KTGAIR-EF 892  
Qy 689 CGTTWVLNDDOHQY---GN-WQAI 710  
Db 893 SGDTEYLLNDSREKYSFKXGNWNGV 918

RESULT 15  
US-11-052-554A-283  
; Sequence 283, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 283  
; LENGTH: 1461  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-283

Query Match 3.5%; Score 140.5; DB 7; Length 1461;  
Best Local Similarity 18.4%; Pred. No. 0.056;  
Matches 148; Conservative 107; Mismatches 304; Indels 245; Gaps 37;  
Qy 15 GGGFVDGIVFNEGA-----PGILYVRT-DIGGMYRWDANGRWIPLLDWGVNNGVNG 67  
Db 515 GTHVLVTGLADAGSNVVALDPAV--TRTLDRO-----ANTIFVTDTAAGNTGAASRA 565  
Qy 68 VVSTAADPINTNKVAAVGMVYNSWDPNDGATLRSDDGATWQITPLPKLGNMFGRM 127  
Db 566 ITLGVVSPLIT-----INTVSGD--IISAEKGA-----PFLTGSTQOAE 606  
Qy 128 GERLAVDPNNDNIIYFGAPSGKGLWRSTDSGATWS-----QMTNFPDVGTVI-ANPTDT 180  
Db 607 GQTVTV-----TLAQSFITTVQADSGSWSLTPAAAMGNLPDCAVAITASVTDL 655

```
QY 181 TGYQSDIQVWVAFDKSSSLGQASKTI FGVADPNN- PVFWSRD-----GG 227
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
656 SGNMGNTSRITITVDSQAPALSIDLPTADNTINAAESQDLPI TGTDAQPGQVTVTLNG 715
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228 ATQOAVPGAPTGFPHKGVDPVNHVLYIATSN TGPGYDGS SDVWKFSVTS GTWTRISP 287
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
716 QTYQ-----GVQPGTWS-----VTVPAANV GALADGNATV TASVNDVAGNPSSVR 763
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
288 VPSTDTANDYFGXGLTIDR-----QHPNTIMVATQISWMPDTIIFRST DGGATWTRIMD 342
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
764 VALVDATPPVVVTINPVATDNTVINTP EHAQAQIISGTVTGAQAGDI VTVTLNNVDYTTVVD 823
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
343 -----WT-----SY-----NRSRYVLDISAE PWLFGVQPNP 371
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
824 GSGNWSLGVPA SVVSGIADSGSY PVSVSVTDKAGNTGSQLT VTVNTAA----- 871
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
372 PVFSPKLGWM-----DEAMAI DPFNSDRMLYGTG-----ATLYATNDLTKWDSGQI 418
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
872 -----PLIGINSIAGDDVINASEKADLQITGTS DQPVNTAITVTLNGQNTYTTTIDASG-- 925
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
419 HIAPMWKLEETA VNDLISPPSGAPLISALGDLGGFTHADVTA VPSTIFTSPVFTTGTSV 478
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
926 -----NWSVTVPA-----SAVTALQOANYTVTA AVTSDIGNSATASHNVLV 966
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
479 DYAEI NPSIIVRAGSPDPSPQNDRHVAFST DGGKNWFQSGEPGVT TGGTVAA SADSQR 538
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
967 DSA--LPGVTINPVATDDIINAAEAGVAQTISGQVTGAEDGDTVTITLGGNTV TATVGSN 1024
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
539 FVNAPGDGQPVVYVGVFGNSWAASQGV PANAQIRSDRVNPKTFYALNGTFFYRST DGGV 598
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1025 LTWSVDVPAAD-IOALGNGD-----LTYNASVTN-----ONG-----NTGSGT 1061
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
599 TFQPVAAGLP-----SSGAVGVMPH VPGKEGDLWLAASSGLYHST----- 639
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1062 RDITIDANLPGLAVDTVAGDDVNIIEHG-----QALVVTGSSGLAESTPLTVTINNVEY 1117
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
640 -----NGSSWSAITGVSSA-----VNVGF-GKSAPGSSY-----PAVFVV 674
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1118 TTAVQADGWS--VGVTAQVSAWPAGT VNIASVSGESSAGNSVSI THPVTVDLT PPAITI 1175
QY   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
675 GTIGG---VTGAYRSDD---CGTT 692
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1176 NTIATDDVINAEKAGADLTLSGTT 1199
```

Search completed: March 2, 2006, 14:28:13  
Job time : 20.0053 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006, Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:06:58 ; Search time 266.937 Seconds  
(without alignments)  
4927.737 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNVAIGGGGFVD.....YGTGNGRIGVYDGGAPSG 740

Scoring table:

|                           |  |
|---------------------------|--|
| BLOSUM62                  |  |
| Xgapop 10.0 , Xgapext 0.5 |  |
| Ygapop 10.0 , Ygapext 0.5 |  |
| Fgapop 6.0 , Fgapext 7.0  |  |
| Delop 6.0 , Delext 7.0    |  |

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/abs/ABSSWEB.spool/US0917376/runat\_02032006\_091454\_8101/app\_query.fasta\_1  
-DB=Issued Patents NA -QFWT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs806h -USER=US0917376@cgn\_1\_1\_278@runat\_02032006\_091454\_8101  
-NCPU=6 -ICPU=3 -NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

| Issued Patents NA.*                        |
|--|
| 1: /cgn2_6/ptodata/1/ina/1/COMB.seq.*      |
| 2: /cgn2_6/ptodata/1/ina/5/COMB.seq.*      |
| 3: /cgn2_6/ptodata/1/ina/6A/COMB.seq.*     |
| 4: /cgn2_6/ptodata/1/ina/6B/COMB.seq.*     |
| 5: /cgn2_6/ptodata/1/ina/H/COMB.seq.*      |
| 6: /cgn2_6/ptodata/1/ina/PC/TUS/COMB.seq.* |
| 7: /cgn2_6/ptodata/1/ina/PP/COMB.seq.*     |
| 8: /cgn2_6/ptodata/1/ina/RE/COMB.seq.*     |
| 9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length DB | ID                     | Description       |
|------------|-------|-------------|-----------|------------------------|-------------------|
| 1          | 392.5 | 9.7         | 1103      | 3 US-09-533-559-7511   | Sequence 7511, Ap |
| c          | 227.5 | 5.6         | 3129      | 3 US-09-252-991A-13873 | Sequence 13873, A |
| 3          | 227.5 | 5.6         | 8211      | 3 US-09-252-991A-13656 | Sequence 13656, A |
| c          | 203   | 5.0         | 4188      | 3 US-09-252-991A-13774 | Sequence 13774, A |
| c          | 201.5 | 5.0         | 4131      | 3 US-09-252-991A-13773 | Sequence 13773, A |
| c          | 192   | 4.8         | 4411529   | 3 US-09-103-840A-1     | Sequence 1, Appli |
| c          | 184   | 4.6         | 10317     | 3 US-09-902-540-1027   | Sequence 1027, Ap |
| c          | 184   | 4.6         | 4403765   | 3 US-09-103-840A-2     | Sequence 2, Appli |
| 9          | 181.5 | 4.5         | 11679     | 3 US-09-328-352-1377   | Sequence 1377, Ap |

|    |       |     |         |   |                      |                    |
|----|-------|-----|---------|---|----------------------|--------------------|
| 10 | 177   | 4.4 | 29927   | 3 | US-09-949-016-11814  | Sequence 11814, A  |
| 11 | 177   | 4.4 | 29927   | 3 | US-09-949-016-11814  | Sequence 17474, A  |
| 12 | 177   | 4.4 | 29927   | 3 | US-09-949-016-11814  | Sequence 17474, A  |
| 13 | 177   | 4.4 | 37802   | 3 | US-09-949-016-12639  | Sequence 12639, A  |
| 14 | 175.5 | 4.3 | 4287    | 3 | US-09-902-540-2590   | Sequence 2590, Ap  |
| 15 | 175.5 | 4.3 | 16584   | 3 | US-09-902-540-1119   | Sequence 1119, Ap  |
| 16 | 170.5 | 4.2 | 25165   | 3 | US-09-453-702B-39    | Sequence 39, Appl  |
| 17 | 170.5 | 4.2 | 25165   | 3 | US-10-114-170-39     | Sequence 39, Appl  |
| 18 | 170   | 4.2 | 4403765 | 3 | US-09-103-840A-2     | Sequence 2, Appli  |
| 19 | 169   | 4.2 | 4411529 | 3 | US-09-103-840A-1     | Sequence 1, Appli  |
| 20 | 168.5 | 4.2 | 20113   | 3 | US-09-902-540-1173   | Sequence 1173, Ap  |
| 21 | 167.5 | 4.2 | 4236    | 3 | US-09-902-540-5367   | Sequence 5367, Ap  |
| 22 | 167.5 | 4.2 | 34662   | 3 | US-09-902-540-1261   | Sequence 1261, Ap  |
| c  | 164   | 4.1 | 2319    | 3 | US-09-252-991A-13875 | Sequence 13875, A  |
| c  | 163   | 4.0 | 5741    | 2 | US-07-706-699-4      | Sequence 4, Appli  |
| 24 | 163   | 4.0 | 5741    | 2 | US-07-998-931-4      | Sequence 1, Appli  |
| 25 | 163   | 4.0 | 5741    | 2 | US-09-679-279-1      | Sequence 1, Appli  |
| 26 | 160.5 | 4.0 | 47981   | 3 | US-09-248-796A-20    | Sequence 20, Appl  |
| 27 | 159.5 | 4.0 | 4647    | 3 | US-09-252-991A-5730  | Sequence 5730, Ap  |
| 28 | 159.5 | 4.0 | 10419   | 3 | US-09-408-020-3      | Sequence 3, Appli  |
| 29 | 159.5 | 4.0 | 42432   | 3 | US-09-408-020-2      | Sequence 2, Appli  |
| 30 | 159.5 | 3.9 | 13805   | 3 | US-09-902-540-1083   | Sequence 1083, Ap  |
| c  | 159   | 3.9 | 2067    | 3 | US-09-489-039A-2642  | Sequence 2642, Ap  |
| 32 | 157.5 | 3.9 | 3155    | 3 | US-09-710-279-3881   | Sequence 3881, Ap  |
| 33 | 157   | 3.9 | 3155    | 3 | US-09-710-279-3881   | Sequence 9781, Ap  |
| c  | 156.5 | 3.9 | 1860    | 3 | US-09-902-540-597    | Sequence 597, App  |
| c  | 156.5 | 3.9 | 5132    | 3 | US-09-902-540-1228   | Sequence 1228, Ap  |
| 36 | 156   | 3.9 | 26930   | 3 | US-09-902-540-1273   | Sequence 1273, Ap  |
| 37 | 156   | 3.9 | 72704   | 3 | US-09-710-279-3107   | Sequence 3107, Ap  |
| c  | 155.5 | 3.9 | 2733    | 3 | US-08-418-782-1      | Sequence 1, Appli  |
| 39 | 155   | 3.8 | 2235    | 2 | US-08-228-662-1      | Sequence 1, Appli  |
| 40 | 155   | 3.8 | 2235    | 2 | US-08-852-219-1      | Sequence 1, Appli  |
| 41 | 155   | 3.8 | 2235    | 2 | US-09-023-655-1420   | Sequence 1420, Ap  |
| 42 | 154   | 3.8 | 3900    | 3 | US-09-949-016-12988  | Sequence 12988, A  |
| c  | 154   | 3.8 | 16013   | 3 | US-09-710-279-3661   | Sequence 3661, Ap  |
| 44 | 153   | 3.8 | 2625    | 9 | 5457037-4            | Patent No. 5457037 |
| 45 | 153   | 3.8 | 3050    | 3 | US-09-710-279-3661   | Sequence 3661, Ap  |

#### ALIGNMENTS

RESULT 1  
US-09-533-559-7511  
; Sequence 7511, Application US/09533559  
; Patent No. 6902887  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster  
; APPLICANT: Sakari Kauppinen  
; APPLICANT: Ib Groth Clausen  
; APPLICANT: Peter Bjarke Olsen  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE OF INVENTION: Expression  
; FILE REFERENCE: 5849.200-US  
; CURRENT APPLICATION NUMBER: US/09/533,559  
; CURRENT FILING DATE: 2000-03-22  
; EARLIER APPLICATION NUMBER: 09/273,623  
; EARLIER FILING DATE: 1999-03-22  
; NUMBER OF SEQ ID NOS: 7860  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7511  
; LENGTH: 1103  
; TYPE: DNA  
; ORGANISM: Tricoderma reesei  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(1103)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-533-559-7511  
Alignment Scores: 1.47e-25 Length: 1103  
Pred. No.: 392.50 Matches: 114  
Score:

|                        |       |               |     |
|------------------------|-------|---------------|-----|
| Percent Similarity:    | 45.0% | Conservative: | 36  |
| Best Local Similarity: | 34.2% | Mismatches:   | 125 |
| Query Match:           | 9.7%  | Indels:       | 58  |
| DB:                    | 3     | Gaps:         | 10  |

US-09-917-376-3 (1-740) x US-09-533-559-7511 (1-1103)

|    |     |   |     |
|----|-----|---|-----|
| QY | 302 | GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp    | 321 |
| Db | 3   | GGCGTTGGCTCGATTTCACAAAGCCAGGAACCTTGTTGTTGCTTCTTTGAACCTCTTGG     | 62  |
| QY | 322 | TrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp    | 341 |
| Db | 63  | TGGCCAGATGCTCAGCTGTTTCGTCGACCGACTCTGGGCAACATCGAGCCGCGATCTGG     | 122 |
| QY | 342 | AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro    | 361 |
| Db | 123 | GGTGGGGCGAGTATCCAGCTGACACCTATTACTACAGCACTCAACTCCCAAGCACCG       | 182 |
| QY | 362 | TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro           | 376 |
| Db | 183 | TGGATCAAGAAACAACTTTTATCATGATGTACGAGCGAGTCACCGTCCGATGCTCTNATCAAG | 242 |
| QY | 377 | LysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu    | 396 |
| Db | 243 | CGCTCGGCTCGATGATGATGACTCTNTCGAGATTGACCCACCCGACAGCAANNACTGCGTT   | 302 |
| QY | 397 | TyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuTyrLysTrpAspSer----      | 414 |
| Db | 303 | TTACGGCACCGGATGACAAATNTTGGCGGCACGATTCACCACTGGACACGGCGCC         | 362 |
| QY | 415 | -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluGluThrAlaVal    | 432 |
| Db | 363 | ACAATGTGGTCAATCCAAATTACTTGGCAGCGGATTCGAAGAAATT-----TTTCCGT      | 416 |
| QY | 432 | AlaAsnAspLeuIleSerProProSerGly-----                             | 441 |
| Db | 417 | TCAAGGACCTGGCCCTTTTACCCGGGGGGAAGCGAGCTTTTGCCGCCAAGTCGGGACG      | 476 |
| QY | 442 | -----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV       | 459 |
| Db | 477 | ANCAACGGGTTTACGTTTTGGCCGAGAAACACGACCTTGG-----                   | 517 |
| QY | 459 | AlThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr-SerVal    | 478 |
| Db | 518 | -----ACATTCGCGCAGAGGTTTGGGCAACTCCCATGGGCCACTCGACGAGCGTC         | 572 |
| QY | 479 | AspTyrAla--GluLeuAsnProSerIleIleValArgAlaGlySer-----            | 493 |
| Db | 573 | GACTACGCGGGAACCTCGGTCAAGAGCGTTCGTCCGCGTCGGCAACACCGCGCGGACG      | 632 |
| QY | 494 | -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS                | 508 |
| Db | 633 | CAACMAGTGCCCATTTTTTCCGACGCGCGCGCGCGAGCTNGGAA-----GCAATTC        | 686 |
| QY | 508 | erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG    | 528 |
| Db | 687 | GAACTAACGCT-----GGNTCCGAAACACCGTTTCCATTGAAACGGCG                | 728 |
| QY | 528 | lyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG    | 548 |
| Db | 729 | CGCGGCTGGCTTATTTCGCGCGACGCGCACACGATCTCTGTGTCGACCGCTCGTCGGC-     | 787 |
| QY | 548 | lnProValValTyrAlaValGlyPheClyAsnSerTrpAlaAlaSerGlnGlyValProA    | 568 |
| Db | 788 | -----GTGACGGCTCGCAGTTCCAGGGCAGCTTTCCTCCGTCGACGCTCGCGCG          | 839 |
| QY | 568 | laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnG    | 588 |
| Db | 840 | CGGGCGCGTTCATCGNCTCGGACAAAGAGACCAACAGCGTNTTCTACGCGCGCTCGGAT     | 899 |
| QY | 588 | lyThrPheTyrArgSerThrAspGlyGly                                   | 597 |

Db 900 CGACCTTTTACGTACGACGACACCGGC 928

RESULT 2

US-09-252-991A-13873/c

; Sequence 13873, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13873

; LENGTH: 3129

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13873

Alignment Scores:

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Pred. No.:             | 4,22e-10 | Length:       | 3129 |
| Score:                 | 227.50   | Matches:      | 218  |
| Percent Similarity:    | 31.3%    | Conservative: | 90   |
| Best Local Similarity: | 22.2%    | Mismatches:   | 360  |
| Query Match:           | 5.6%     | Indels:       | 319  |
| DB:                    | 3        | Gaps:         | 52   |

US-09-917-376-3 (1-740) x US-09-252-991A-13873 (1-3129)

|    |      |  |                                   |        |    |
|----|------|--|-----------------------------------|--------|----|
| Qy | 20   | AspGlyIleValPheAsnGluGlyAla----                                | ProGlyIleLeuTyrValArgThrAspIle    | 38     |    |
| Db | 2993 | AACGGCGTGTCTATCAGCGGCACCGCGCGGTGCCACCGTGCACCGTCCACCGATGCC      | 2934                              |        |    |
| Qy | 39   | GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal   | 58                                |        |    |
| Db | 2933 | GGC-----GGCAACCGATAGGCGAGGTACCGCGCAG                           | 2901                              |        |    |
| Qy | 59   | GlyTrpAsnAsnTrpGlyTyr-----                                     | AsnGly-----                       | ValVal | 69 |
| Db | 2900 | GGCAGCGGCACTGGAGCTTCACCGCGGCACGCCCGCCGCAACGCGCAGCGTGTGATCGTC   | 2841                              |        |    |
| Qy | 70   | SerIleAlaAlaAspProIle-----                                     | AsnThrAsnLysValTrpAla-----        | Ala    | 84 |
| Db | 2840 | GCCACGCGCACCCAGCCGACCGGCAATACCGGCCCGCAGCGCCGCCACCGTGTGACGCG    | 2781                              |        |    |
| Qy | 85   | ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu-----          | 100                               |        |    |
| Db | 2780 | GTGCGCGCGCGCGCGGTGATCGATCCGAGCAACGCGCAGCAGCATCAGCGGCACCGCG     | 2721                              |        |    |
| Qy | 101  | -----  | ArgSer                            | 102    |    |
| Db | 2720 | GAGCGCGGCGCAAGGTGATCTCACCGAGGCAACCGCAACCCGATCGCGCAACCCACC      | 2661                              |        |    |
| Qy | 103  | SerAspGlnGlyAlaThrTrpGlnIleThrPro---                           | LeuProPheLysLeuGly-----           | 119    |    |
| Db | 2660 | GCCGAGCGGACGCGGCAACTGACCTTCAGCGCCGCCACGCCGCTGGCCCAACGCGCAGGTG  | 2601                              |        |    |
| Qy | 120  | -----  | GlyAsnMetProGlyArgGlyMetGlyGluArg | 130    |    |
| Db | 2600 | GTCAACCGCGTGCCCGAGGACCTTCGCGGCAATACCGGCCCGCAGGAGGACACTACCGGTG  | 2541                              |        |    |
| Qy | 131  | LeuAlaValAspProAsn-----  | AsnAspAsnIleLeuTyrPhe             | 143    |    |
| Db | 2540 | GAGCGCGTGGCGCGAACACGCGCTGTGTTCAATCCGAGCAACGGCAACCTGTCTACACGT   | 2481                              |        |    |
| Qy | 144  | GlyAlaProSerGlyLysGlyLeu-----                                  | Trp                               | 152    |    |
| Db | 2480 | ACCGCGGCGCGGCGGACGACCGGTGACCTTGACCGAGCGCAACCGGCAACCGGATCGGCCAG | 2421                              |        |    |



|    |      |  |      |
|----|------|--|------|
| QY | 153  | ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe             | 167  |
| Db | 2420 | ACCACCGCGGATGGCAGCGCAACTGGAGCTTACGCCCGCGTCCCAACTACCCAAAC       | 2364 |
| QY | 168  | ProAspValGlyThrTyrIle-----AlaAsnProThrAspThrThrGlyTyrGlnSerAsp | 186  |
| Db | 2363 | -----GGCACCCTGGTCAACGTGACCGGAGCGAGCGCGCGCAATACACG              | 2316 |
| QY | 187  | IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln-----    | 204  |
| Db | 2315 | -----GCTCCCGTACCAACGAGCGTGGATTCTCTCGCTCGCTCGATCCCGAGGTGGAT     | 2262 |
| QY | 205  | AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe-----       | 221  |
| Db | 2261 | CCGAGCAACGCTTCGGTGATCAGCGGCACCGCGGACCGCGCAACACCATCATCATCACC    | 2202 |
| QY | 222  | -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla                         | 232  |
| Db | 2201 | GATGGCAACGGCAACCGGATTGGCCAGGTTCACCGCGGACGCGGTAACCTGTCCTTC      | 2142 |
| QY | 233  | ValProGlyAlaPro-----   | 237  |
| Db | 2141 | ACTCCAGGCATCCCGCTGCCGATGGACGCTGGTCAACGTGGTGGCGGAGCCCAAGC       | 2082 |
| QY | 238  | -----ThrGlyPheIleProHisLysGlyVal                               | 246  |
| Db | 2081 | AATGTCGACAGTGCAGCGCGGTGATCACTGTGGATGGCGTGGCCCGCGCGCGGTG        | 2022 |
| QY | 247  | PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr----  | 265  |
| Db | 2021 | ATCGATCCGAGCAACGGCACCGAGATAGCGGTACCGCGAGGCGCGCGCGCGTATC        | 1962 |
| QY | 266  | -----AspGlySerSerGlyAspValTrpLysPheSerValThr-----SerGlyThrTrp  | 282  |
| Db | 1961 | CTCACCAGTGGCGCGGCAACCCGATCGCGCAGGCCACCGCGGAGCGGCGGCGCAACTGG    | 1902 |
| QY | 283  | ThrArgIleSerProValPro-----SerThrAsp                            | 292  |
| Db | 1901 | ACGTTACCCCGGCGCACCCCGCTGGCCCAACCGCACCGTGTCAACGCCGTGGCCCGCAGC   | 1842 |
| QY | 293  | ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro-----    | 310  |
| Db | 1841 | CGCGCGGCAATACAGCGGTCCCGCGCAGCTCACCGTGCATGCCATCGCCCGCGCGC       | 1782 |
| QY | 311  | -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp                | 324  |
| Db | 1781 | CCGGTGATCAATCCGAGCAACGAGTGTCTATCAGCGGTACGCGGAACCGCGGCCACG      | 1722 |
| QY | 325  | ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr                | 338  |
| Db | 1721 | GTGATCTCTC-----ACCGAGCGCAACCGCAACCGCATCGCGGAGTCCCGCCGAC        | 1671 |
| QY | 339  | ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerLe                   | 351  |
| Db | 1670 | GSCAGCGCAAGTGGCTTTTACGCCCGCGCACCGCTGGCCCAATGGCAGCGTATCAAT      | 1611 |
| QY | 351  | uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr   | 371  |
| Db | 1610 | GCGCTGG-----CCGAGGACCGCGCGCGCAACACAGCAGTCTC-----CC             | 1572 |
| QY | 371  | oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs   | 391  |
| Db | 1571 | ACGAGCGCCA-----CCGTGCTCGCTGGCGGCAGCA-GCCCGGTGATCGATCCGAGCAA    | 1516 |
| QY | 391  | nSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAlaThrAsnAs       | 408  |
| Db | 1515 | CGGTAGC---GTGATCGCGGTACCGCGAGGCTGGTGCCACGCTGATCTCTCACC-----    | 1464 |
| QY | 408  | pLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeuGln     | 428  |
| Db | 1463 | -----GACGCAACGGCAAC-----CCGATCGCGCCAGGTCCACCGC                 | 1429 |
| QY | 428  | uGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeu-----            | 444  |
| Db | 1428 | CGATGGCAGCGCAACTGGAGCTTACGCC-----GGCACCCTCGTCCCAATGGCAC        | 1375 |
| QY | 445  | -----IleSerAlaLeuGly-----AspLeuGlyGlyPheThrHisAlaAspVal---     | 461  |
| Db | 1374 | GTTGGTCAATGGTGGCGCCAGAGCTGCCGGCAACACGAGCGCGCGCCAGCACAC         | 1315 |
| QY | 461  | aValProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal--    | 478  |
| Db | 1314 | GTTGACTCGTGGCGCGCGCGCGCGGTGATCGACCGCAACGAGCGAGCGGTGATCGC       | 1255 |
| QY | 479  | AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe    | 498  |
| Db | 1254 | CGTACCGCGGAAGCGGTGCGAGCTGATCTTACCGATGGCGGCGGCAACCGATCGG        | 1195 |
| QY | 498  | rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----      | 515  |
| Db | 1194 | CCAGGCC-----ACCGCGATGGCAGCGCAACTGGAGCTTTCAC                    | 1156 |
| QY | 516  | -----PheGlnGlySerGluProGlyGln                                  | 523  |
| Db | 1155 | CCCGGCGCAGCGTGGCCCAACGCGCACGCTGATCAATGCGGTGGCCCGCAGGATCGGCGG   | 1096 |
| QY | 523  | yValThrThrGlyGlyThr-----                                       | 529  |
| Db | 1095 | CAATACGAGCGCGCACCGACGACGCGGTGACGCGGTGGCCCGCCCGCCCGCGTGT        | 1036 |
| QY | 530  | -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl                    | 542  |
| Db | 1035 | CAACCCGAGCAACGCGAGCGTGTATCGCGGTACCGGGAAGCGCGCCCGCGTGTCT        | 976  |
| QY | 542  | aProGlyAspProGlyGlnProVal-----ValTyrAlaValGlyPheGlyAsnSerTr    | 560  |
| Db | 975  | CACCGAGCGCGCGCAACCCGATCGCGCAGGTACCGCGCGGCGGCGGCGGCAAC---       | 919  |
| QY | 560  | pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer-----            | 574  |
| Db | 918  | GAGCTTACGCCCGCGCACCGCGCTGGCCCAACCGCTCGGTGATCAATCGGTGGCCCGCAG   | 859  |
| QY | 575  | -----AspArgValAsnProLysTh                                      | 581  |
| Db | 858  | CGCCGCGCAACACGAGCGCGCGCCAGCACCGAGTGGTGGTGGTGGTGGTGGTGGTGGT     | 799  |
| QY | 581  | r---PheTyrAlaLeuSerAsnGlyThrPheTyrArgSer-----                  | 593  |
| Db | 798  | CCCGTGTCTGATCCGAGCAACGCTACGCTGATCAGCGGTACGCGGAGCGCGGCGCAC      | 739  |
| QY | 594  | -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS        | 610  |
| Db | 738  | GTTGATCTTACCGAGCGGCGC-----GGCAACCGGATACGCGGCGCACCGCGGATGG      | 685  |
| QY | 610  | erGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuA   | 630  |
| Db | 684  | CAGCGGCAACTGGAGCTTCACTCCCGGCACA-CCGTGACCAACGCGCGCGTATCAATG     | 626  |
| QY | 630  | laAlaSerSerGlyLeuTyrHisSerThrAsnGly-----GlySerSerTrpSerAlaI    | 648  |
| Db | 625  | CGTGGCGCCAGCAGCGCGCGGCAACACGAGCGGTCCGCTCAGCACCATGATGGAGCGG     | 566  |
| QY | 648  | leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerT   | 668  |
| Db | 565  | TGGCCCGCGCACCCCGGTGATCGACCCGAGCAATGGTGTCAAACTCAGCGGCGCACCGCG   | 506  |
| QY | 668  | yrProAlaValPheValGlyThrIleGlyGly-----                          | 679  |
| Db | 505  | AACCGCGGTCCGGGTGATCTTCAACCGATGGCAATGGCAACCGATCGCGCCAGCTCG      | 446  |
| QY | 680  | -----  | 681  |
| Db | 445  | CCGACGCTAGCGGCAACTGGACCTTTCACACCGCGGCGACCGCTGGCGCAACGCGCGGTG   | 386  |
| QY | 681  | hrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGlnH      | 701  |

```
Db 385 TCACGCGGTGGCCAGGACCGCGCGCAATACCAG-----CGTCCGGCCAGC 336
Qy 701 isGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg-Val 720
Db 335 ACCACGGTGATACGGTGGCGCGCCACCGCGTGATCAATGCCAGCAACGCGAGGTG 276
Qy 721 TyrIleGlyThrAsn-----GlyArgGlyLeuValTyrGlyAspIleGlyGlyAlaPro 738
Db 275 ATCACCAGGACCGCGAGGTGGCGGCAAAAGTGATCTCACCGACGGAACGCGCAACCCG 216
Qy 739 SerGly 740
Db 215 ATCCGC 210

RESULT 3
US-09-252-991A-13656
; Sequence 13656, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13656
; LENGTH: 8211
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13656

Alignment Scores:
Pred. No.: 1,84e-09 Length: 8211
Score: 227.50 Matches: 218
Percent Similarity: 31.3% Conservative: 90
Best Local Similarity: 22.2% Mismatches: 360
Query Match: 5.6% Indels: 319
DB: 3 Gaps: 52

US-09-917-376-3 (1-740) x US-09-252-991A-13656 (1-8211)
Qy 20 AspGlyIleValPheAsnGluGlyAla---ProGlyIleLeuTyrValArgThrAspIle 38
Db 3574 AACGGCGTGCTCATCAGCGGCACCGCGCGGTGCCACCGTGCACCCCTCACCGATGCC 3633
Qy 39 GlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal 58
Db 3634 GGC-----GGCAACCGGATAGGCGAGGTGCACCGCGCAG 3666
Qy 59 GlyTrpAsnAsnTrpGlyTyr-----AsnGly-----ValVal 69
Db 3667 GGCAGCGGCACTGGAGCTTCACCGCGGCACCGCGCGCGCCCAACGCGCGTGATGCTC 3726
Qy 70 SerIleAlaAlaAspProIle---AsnThrAsnLysValTrpAla-----Ala 84
Db 3727 GCCACGCGCACCGACCGACCGGCAATACCGCGCGCGCGCGCCACCGCGTGAGCGG 3786
Qy 85 ValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu-----100
Db 3787 GTGGCGCGCGCGCGCGGTGATCGATCCGACCAACGCGCACGACCATCAGCGGACCGCG 3846
Qy 101 -----ArgSer 102
Db 3847 GAGCGCGCGCGCGNAGTGATCTCACCGACGCGCAACCGGCAACCCGATCGCGGAAACCA 3906
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly-----119
Db 103 SerAspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGly-----119
```

```
Db 3907 GCCACGCGCAGCGCAACTGGACCTTACGGCCGCGCCACCGCCCTGGCCACACGCGAGGTG 3966
Qy 120 -----GlyAsnMetProGlyArgGlyMetGlyGluArg 130
Db 3967 GTCAACGCGGTGGCCAGGACCTTGGGGCAATACCGGCGCGCGGAGCAGCACTACCGTG 4026
Qy 131 LeuAlaValAspProAsn-----AsnAspAsnIleLeuTyrPhe 143
Db 4027 GACCGGTGGCGCGCAACACCGCTTGTCAATCCGAGCAACGCGCAACCTGCTCAACGCT 4086
Qy 144 GlyAlaProSerGlyLysGlyLeu-----Trp 152
Db 4087 ACCCGCGAGCGCGCAGCACCGCTTGACCGGACGCGCAACCGGATCGGCCAG 4146
Qy 153 ArgSerThrAspSerGlyAlaThrTrp-----SerGlnMetThrAsnPhe 167
Db 4147 ACCACCGCGGTGGCAGCGGCAACTGGAGCTTCAGCCCGGCTCGCACTACCCACAC---4203
Qy 168 ProAspValGlyThrTyrIle---AlaAsnProThrAspThrThrGlyTyrGlnSerAsp 186
Db 4204 -----GGCACCGTGGTCAACGTGACCGCGAGCGACCGCGCGCAATACCCAGC---4251
Qy 187 IleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln-----204
Db 4252 -----GCTCCCGCTACACGACGCGTGGATTCCTCGCTGCGCTGCATCCCGAGGTGGAT 4305
Qy 205 ---AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPhe-----221
Db 4306 CCGAGCAACGGTTCGGTGATCAGCGGCACCGCGGACCGCGGCAACACCATCATCATCACC 4365
Qy 222 -----TrpSerArgAspGlyGlyAlaThrTrpGlnAla 232
Db 4366 GATGGCAACGCGCAACCGGATGGCCAGGTACCGCGCAGCGCGGTAACCTGGTCTTC 4425
Qy 233 ValProGlyAlaPro-----237
Db 4426 ACTCCAGGCATCCCGCTGCGCGATGGCACGGTGTCAACGTGGTGGCGCGCAGCCCAAGC 4485
Qy 238 -----ThrGlyPheIleProHisLysGlyVal 246
Db 4486 AATGTCACAGTGGCGCGCGGTGATCATTGTGGATGGTGGCCCGCGCGCGCGGTG 4545
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr---265
Db 4546 ATCGATCCGAGCAACGCGCACCGAGATAGCGGTACCGCGAGCGCGCGCGCGGTGATC 4605
Qy 266 -----AspGlySerSerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp 282
Db 4606 CTCACCGATGGCGCGCGCAACCGATCGCGCAGCGCACCGCGCAGCGCGCAACTGG 4665
Qy 283 ThrArgIleSerProValPro-----SerThrAsp 292
Db 4666 ACGTTCACCGCGCGCACCGCGTGGCCACCGCGACCGGTGATCAACCGCGTGGCCAGGAC 4725
Qy 293 ThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisPro-----310
Db 4726 CCGCGCGCAATACCGAGGTCCCGCGCAGCGGTACCGTGCATGCCATGCCCGCGCGCGG 4785
Qy 311 -----AsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324
Db 4786 CCGGTGATCAATCCGAGCAACGGAGTCTGTCATCAGCGGTACGCGGAAAGCGCGGCGCACG 4845
Qy 325 ThrIleIlePheArgSerThrAspGlyGly-----AlaThrTrpThr 338
Db 4846 GTGATCTCTC-----ACCGAGCGCAACCGCAACCGCATCGCGCAGGTACACCGCCAGC 4896
Qy 339 ArgIleTrpAspTrpThr-SerTyrProAsnArg-----SerLe 351
Db 4897 GGCAGCGCGAAGTGGGCTTTCACCGCGCGCACCGCGTGGCGCAATGGCAGCGGTATCAAT 4956
Qy 351 uArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPr 371
Db 4957 GCGGTGG-----CCGAGCGCGCGCGCAACACGAGCAGTC-----CC 4995
```

|    |      |   |      |
|----|------|---|------|
| Qy | 371  | oProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAs  | 391  |
| Db | 4996 | ACCAAGCCCA---CCGTCGACTCGTGGCCAGCA-GCCCGGTGATCGATCCGAGCAA      | 5051 |
| Qy | 391  | nSerAspArgMetLeuTyRglyThr-----GlyAlaThrLeuTyAlaThrAsnAs       | 408  |
| Db | 5052 | CGGTAGC---GTGATCGCGGTACCGCCGAGGCTGGTGCACGCTGATCCTCAC          | 5103 |
| Qy | 408  | pLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGl  | 428  |
| Db | 5104 | -----GACGGCAACGGCAAC-----CCGATCGGCAGGTCAACCGC                 | 5138 |
| Qy | 428  | uGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeu-----        | 444  |
| Db | 5139 | CGATGGCAGCGGCAACTGGAGCTTCAGCCC-----GGCAGCCGCTGTCCAATGGCAC     | 5192 |
| Qy | 445  | IleSerAlaLeuGly---AspLeuGlyGlyPheThrHisAlaAspVal---ThrAl      | 461  |
| Db | 5193 | GGTGGTCAATGCGGTGGCCAGGACGCTGCCGCAACACCAGCGGCCGCCAGCACCCAC     | 5252 |
| Qy | 461  | avalProSerThrIlePheThrSerProValPhe-----ThrThrGlyThrSerVal--   | 478  |
| Db | 5253 | GGTGGACTCGTGGCGCGCGCGCGCGGTGATCGACCCGAGCAACGCGAGCGTGATCGC     | 5312 |
| Qy | 479  | AspTyAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe    | 498  |
| Db | 5313 | CGGTACCGCGGAAGCCGTCGACGGTGTCTCCACCGATGGCGGCGCAACCCGATCGG      | 5372 |
| Qy | 498  | rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrp-----     | 515  |
| Db | 5373 | CCAGGCC-----ACCGCGGATGGCAGCGGCACTGGAGCTTCAC                   | 5411 |
| Qy | 516  | -----PheGlnGlySerGluProGlyGl                                  | 523  |
| Db | 5412 | CCCGGGCACGCGTGCCCAACGGCAGCGGTGATCAATCGGTGGCCCGAGATCCGGCCGG    | 5471 |
| Qy | 523  | yValThrThrGlyGlyThr-----                                      | 529  |
| Db | 5472 | CAATACGAGCGGCCACCAAGCACACCGGTGGACGGGTGGCCCGCCGCCACCCCGGTGT    | 5531 |
| Qy | 530  | -----ValAlaAlaSerAlaAspGlySerArgPheValTrpAl                   | 542  |
| Db | 5532 | CAACCCGAGCAACGGCAGCGTGTATCGCGGTACCGCGGAAGCCGCGCCACGCGTGATCT   | 5591 |
| Qy | 542  | aProGlyAspProGlyGlnProVal-----ValTyAlaValGlyPheGlyAsnSerTr    | 560  |
| Db | 5592 | CACGACGGCGGGCAACCCGATCGCCAGGTACCGCGCGCAGCGCAGCGGCAC           | 5648 |
| Qy | 560  | pAlaAlaSerGlnGlyValPro---AlaAsnAlaGlnIleArgSer-----           | 574  |
| Db | 5649 | GAGTTTCACGCCCGGCACGCGCGGTGGCCACAGCGCTCGGTGATCAATGCGGTGGCCAGGA | 5708 |
| Qy | 575  | -----AspArgValAsnProLysTh                                     | 591  |
| Db | 5709 | CGCGCGCGGCAACACCAGCGGCCCGGCCAGCACACCGGTGATCTGTTAGCCCGGCCAC    | 5768 |
| Qy | 581  | r---PheTyAlaLeuSerAsnGlyThrPheTyArgSer-----                   | 593  |
| Db | 5769 | CCCGGTGCTCGATCCGAGCAACGGTACGGTGATCAGCGGTACCGCGTACCGCGGCCAC    | 5828 |
| Qy | 594  | -----ThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu-ProSerS       | 610  |
| Db | 5829 | GGTGTCTCTCACGACGGCGGC-----GGCAACCCGATACCGCAGGCCACCGCCGATGG    | 5882 |
| Qy | 610  | erGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuA  | 630  |
| Db | 5883 | CAGCGGCAACTGGAGCTTCACTCCGGGCACA-CGGCTGACCAAGCGCACGGTGATCAATG  | 5941 |
| Qy | 630  | laAlaSerSerGlyLeuTyHisSerThrAsnGly-----GlySerSerTrpSerAlaI    | 648  |
| Db | 5942 | CGGTGGCCAGGACGCCCGGCAACACGACGCGTCCGGTCAGCACCAACGATGGACGGG     | 6001 |

|    |      |  |      |
|----|------|--|------|
| Qy | 648  | leThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSert | 668  |
| Db | 6002 | TGGCCCGGCGCACCCCGGTGATGCAGCCGAGCAATGGTGTCAAACTCAGCGGCACGCGC  | 6061 |
| Qy | 668  | yrProAlaValPheValValGlyThrIleGlyGly                          | 679  |
| Db | 6062 | AACCCGCGTCGCGGTGATCCTCACCAGTGGCAATGGCAACCCGATCGGCACGACCTCG   | 6121 |
| Qy | 680  | -----Valt  | 681  |
| Db | 6122 | CCGACGGTAGCGCAACTGGACCTTCACACCGGGCAGCCGCTGGCCACCGCACGGTGG    | 6181 |
| Qy | 681  | hrGlyAlaTyArGSerAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnH     | 701  |
| Db | 6182 | TCAACGCGGTGGCCAGGACCCGCGCGCAATACCAAG-----CGTCCGGGCAGC        | 6231 |
| Qy | 701  | isGlnTyrglyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArg-Val | 720  |
| Db | 6232 | ACCACGGTGGATACGGTGGCGCGGCACGCCCGGTGATCAATGCCAGCAACGGCAGCGGTG | 6291 |
| Qy | 721  | TyrIleGlyThrAsn-----GlyArGlyIleValTyrglyAspIleGlyGlyAlaPro   | 738  |
| Db | 6292 | ATCACCGGCACCGCGAGGTGCGGGCCAAAGTGATCTCACCGACGGCAACGGCAACCCG   | 6351 |
| Qy | 739  | SerGly 740   |      |
| Db | 6352 | ATCGGC 6357  |      |

RESULT 4

US-09-252-991A-13774/c

; Sequence 13774, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUD

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13774

; LENGTH: 4188

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13774

|                        |                  |
|------------------------|------------------|
| Alignment Scores:      |                  |
| Pred. No.:             | 1,02e-07         |
| Score:                 | 203.00           |
| Percent Similarity:    | 32.4%            |
| Best Local Similarity: | 22.2%            |
| Query Match:           | 5.0%             |
| DB:                    | 3                |
|                        | Length: 4188     |
|                        | Matches: 186     |
|                        | Conservative: 85 |
|                        | Mismatches: 297  |
|                        | Indels: 274      |
|                        | Gaps: 46         |

US-09-917-376-3 (1-740) x US-09-252-991A-13774 (1-4188)

|    |      |   |      |
|----|------|---|------|
| Qy | 47   | AlaAsnGlyArgTrp-----IleProLeuLeuAspTrpValGlyTrpAsn            | 61   |
| Db | 2496 | GCCAACGGCAACTGGTCTTTTCACGCGCTGCACCCCGTCCGGGACG-GTACCGTGGTCA   | 2439 |
| Qy | 62   | AsnTrp-----GlyTyraenGlyValSerIleAlaAlaAspProIleAsnThrAsn      | 79   |
| Db | 2438 | ACGTGGTGGCCAGGGATCGCGCGGCAACAGCAGTCCGCGGCGCAGCGTTA-----       | 2388 |
| Qy | 80   | LysValTrpAlaAlaValGlyMetTyThrAsnSerTrpAspProAsnAspGlyAlaIle   | 99   |
| Db | 2387 | ---CCGTGGAT-GCCGTGGCGCGCGGCACGCCCCCGTGTATCCGAGCAACGATGACGACC  | 2332 |
| Qy | 100  | LeuArg----SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeu | 118  |

Db 2331 CTACGGCGACCGCGGCGCGCTACCGTGACCTGACG----- 2290  
Qy 119 GlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAsp 138  
Db 2289 GACGGCAAC-----GGCAACCCGATTGGCCAGGTACCGGC----- 2254  
Qy 139 AsnIleLeuTyrPheGlyAlaProSerGlyGlyLeuTyrArgSerThrAspSerGly 158  
Db 2253 -----GACGGCAGCGGCAACTGG----- 2236  
Qy 159 AlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIle---AlaAsnPro 177  
Db 2235 ---ACCTTACCCCGACGACCGCGCTGCCCAAC---GGCAGCGTGGTCAACGCCACCGCGCT 2182  
Qy 178 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLys 197  
Db 2181 ACCGACCCGTCGGCAACCGCAGTTCGCGCGCAGCGTCAACCGTGGAGCCCGCGGCACCG 2122  
Qy 198 SerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro--- 216  
Db 2121 GCCACGGCAGTGGTCAACCCCGACGACCGCACACCGTCAACCGCGCAGCGCGCGCGCGC 2062  
Qy 217 -----AsnAsnProValPheTrp---SerArgAspGly 226  
Db 2061 GCCACCGTGACCTGGCGGATGGCAACCGGCAATCCATCGCGGCAAGGTCAACCGCGGATGGC 2002  
Qy 227 GlyAlaThrTrpGlnAlaValProGlyAlaPro----- 237  
Db 2001 AGCGGCAACTGGAGCTTCACTCCGACACCGCGTGTGCCAACGCGACCGCGTGTCAACGCC 1942  
Qy 238 -----ThrGlyPhe 240  
Db 1941 ACGGCCACCGACGCGCTCCGGCAACACACAGTGGCGGCAGCAGTGTCCCGTGGACTCGGTA 1882  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
Db 1881 GCCCGCGCCACCGCAGTGATCAACCCCGACGACCCGACGACCGTCAACCGCGCGCGCGCAG 1822  
Qy 261 ThrGlyGlyProTyr-----AspGlySerSerGlyAspValTyrLysPheSerVal 277  
Db 1821 CCGGGAGCAGCGTGACCTGACCGATGGGCAACGGCAACCGATCGCGCCAGGTCAACCGCC 1762  
Qy 278 Thr---SerGlyThrTrpThrArgIleSerProValPro----- 289  
Db 1761 GACGGCAGCGCACTGGAGCTTCAACCGCTGCACGCGCTGGCGGATGGNACCGTGGTC 1702  
Qy 290 -----SerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAsp 306  
Db 1701 AACGCCACCGCGCCAGTCCGGCGGCAACACACGCGCGCCAGGCGCAGCACCACCGTGCAT 1642  
Qy 307 ArgGlnHisProAsnThrIleMetVal-----AlaThrGlnIleSerTrpTrp 322  
Db 1641 GGGGTGGCGCGCACCGCGCGACCGTCAACCTGTGAGCAACGGCAGCAGCTCAGCGGCACT 1582  
Qy 323 ProAsp-----ThrIleIlePheArg----- 329  
Db 1581 GCGGAACCGGGCAGCAGGTGATCTCACGACGGCAACCGCAATCCGATCGCGCGAGTGC 1522  
Qy 330 SerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArg 349  
Db 1521 ACCGCGACGCGCAGCGCAACTGGACC-----TACACCCCGT---CCACGCGGA 1476  
Qy 350 SerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPro 369  
Db 1475 TCG-----CCAACGGCACCGTGGTCAACGTGGCGCCGACG 1437  
Qy 370 AsnProProValPro-----SerProLysLeuGlyTrpMetAspGluAlaMetAla--- 386  
Db 1436 CCGCGCGCAATAGACGCGCGCGCGCAGCTACCGGTGGAC---TCGCAAGCCCGCGCGCT 1378  
Qy 387 -----IleAspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyVala 401  
Db 392

Db 1377 CCGGTGTCTAACCCCGAGCAACGGC---ACCAGCTCTACGCGGCAACCGCGAGCGCGCGCT 1321  
Qy 402 ThrLeuTyrAlaThrAsn-----AspLeuThrLysTrpAspSer 414  
Db 1320 ACCGTGAGCTTACGCGCAACCGGCAACCGGATTTGCCAGGTCAACGCC---GACGGC 1264  
Qy 415 GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAsp 434  
Db 1263 AGTGGCACTGAGGTTTCAACACCGGCGCAGCGCTGGCCACCGGCAACCGTGTCAACGCC 1204  
Qy 435 LeuIleSerProProSerGly-----AlaProLeuIleSerAlaLeuGlyAspLeu 451  
Db 1203 ACGGCCAGCAGCCGCGCAATACACGCGTCCGCGCAGCACACCGTGCATCGCTGG 1144  
Qy 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
Db 1143 -----GCGCGCGCGCGCTGGTCAATCCGAGCAACGGAGTTC 1105  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1104 GTCATCAGCGGCACG-----GCCGAACCGGCGCCACCGTGCATCGACCGAT 1057  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1056 GGCAGCGCAATCCGATCGGGCAG-----GTCACCGCGCGACGCGC 1018  
Qy 512 GlyLysAsnTrp-----Phe 516  
Db 1017 ACGGGCAACTGAGCTTCAACCGCTGCACGCGGTGGCGGATGGAAACCGTGTCAACGCC 958  
Qy 517 GlnGlySerGluProGlyValThrThrGlyGly-----ThrValAlaAla 532  
Db 957 ACCGTACCGACCGCGCGCGC---AATACCGCGCGCAGGCGCAGCACTACCGTGGACGCC 901  
Qy 533 SerAla-----AspGlySerArgPheValTrpAlaPro 543  
Db 900 ATCGCGCGCGCCACCGCGCAGCGTCAACCTGAGCAATGGCAGCAGC---CTCAGCGCACT 844  
Qy 544 GlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn----- 558  
Db 843 GCGGAACCGGCGCAGCGTGTCTCTACCGCAGCAACCGCAATCCGATCGCGCGAGTGC 784  
Qy 559 -----SerTrp----- 560  
Db 783 ACCGCCGACGCGCGCACTGAGCACTACACCCCGTCCACCGCGATCGCCAACCGTACT 724  
Qy 561 -----AlaIleSerGlnGlyValProAlaAsnAlaGln 571  
Db 723 GTGCTCAACGTGTGGCGGAGGACGCGCGCTTAACAGCAGCGCGCGCGCGCGTGCACC 664  
Qy 572 IleArgSerAspArg-----ValAsnProLysThrPheTyrAlaLeuSer 586  
Db 663 GTCGATTCCAGCGCGCGCGCGCGCGCGTGTATCAACCGC-----AGC 622  
Qy 587 AsnGlyThrPheTyrArgSerThrAspGlyValThrPheGlnProValAlaAlaGly 606  
Db 621 AACGGCGTC-----GTCTACGCGGC 601  
Qy 607 LeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAsp 626  
Db 600 ACCGCCGAGCGCGTGCACCGTGCACCTCACCAGTCCGCGCGCAACCCGATGGGCGAG 541  
Qy 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSer 646  
Db 540 GTC-----ACCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511  
Qy 647 AlaIleThrGlyValSerSerAla-----ValAsnValGlyPheGlyLysSerAla 663  
Db 510 TTCACGCGCGCGCACCGCGCGCGCAACCGGCAACCGGTGATCGTGCACCGCCACCGCGCGCG 451  
Qy 664 ProGlySerSerTyrProAlaValPheValValGlyThrIleGlyValThrGlyAla 683  
Db 450 ACCGGCAATACCG 392



```
Db 2619 GAGCCGCCGCGCAACACGAGCGCCGCGGCGAGCGTGGACTCGGTAGCCCGGCC 2560
Qy SerProValPhe-----ThrThrGlyThrSerValAsp---TyrAlaGluLeuAsnPro 485
Db 2559 ACCCGGTGTCTCGATCCGAGCAACGAGTACCGTATCAGCGGTACCGCGGAGCGCGGCC 2500
Qy SerIlelleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisVal 505
Db 2499 ACGGTATCTCAACGAGCGCGCGGCAACCCGATACCGGAGCC-----2455
Qy 506 AlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThr 525
Db 2454 -----ACCGCCATGTCAGCGGCAACTGG-----AGCTTCACTCCGGGCAACCGCTG 2407
Qy 526 ThrGlyGlyThrVal-----AlaAlaSerAlaAspGlySerArgPheValTrpAlaPro 543
Db 2406 ACCAAGCGCACGATGATCAATGCGGTGGCCGAGGAGCGCCCGCGCAACACGCGGTCCG 2347
Qy 543 -----543
Db 2346 GTCAGCACACAGTGGAGCGGTGGCCCGCCGACCCCGGTGATCGACCGAGCAATGGT 2287
Qy 544 -----GlyAspProGlyGlnProValValTyrAlaValGlyPheGly 557
Db 2286 GTCAAACTCAGCGGCACCGCGCAACCGCGGTCCGGTGATCTCTCACCGATGGCAATGGC 2227
Qy 558 Asn-----SerTrpAlaAlaSerGlnGlyVal 566
Db 2226 AACCGATCGGCAGACCCCTCGCGAGCGGTAGCGGCAACTGGACCTTCACACCGGCGAGC 2167
Qy 567 ProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSer 586
Db 2166 CCG-----CTGGCC 2158
Qy 587 AsnGlyThrPheTyrArg-----SerThrAspGly-----596
Db 2157 AACGGCACGCTGTGTCAACGCGGTGGCCGAGGACCCGCGCGCAATACCGCGGTCCGCC 2098
Qy 597 GlyValThrPheGlnProValAlaAlaGlyLeuPro-----SerSerGlyAla 612
Db 2097 AGCACACGCTGATACGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2038
Qy 613 ValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer 632
Db 2037 GTG---ATCACCGCACCGCGGAGTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1981
Qy 633 SerGlyLeuTyrHisSerThr---AsnGlyGlySerSerTrpSerAlaIleThrGlyVal 651
Db 1980 AACCCGATCGCGAGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1921
Qy 652 -----SerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSer 667
Db 1920 CCGCTCGCCACCGGTACGCGGTATCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1861
Qy 668 TyrProAlaValPheValGlyThrIleGlyGlyValThrGlyAlaValTyrArgSer 686
Db 1860 GGTCCGCGC-----AGCACAC-GGTGGACTCGGTGGCGGCGGCGGCGGCGGCGGCGGCGG 1817
```

## RESULT 6

```
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
```

```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 0.0402 Length: 4411529
Score: 192.00 Matches: 194
Percent Similarity: 35.0% Conservative: 94
Best Local Similarity: 23.6% Mismatches: 323
Query Match: 4.8% Indels: 213
DB: 3 Gaps: 43

US-09-917-376-3 (1-740) x US-09-103-840A-1 (1-4411529)

Qy 10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly-----27
Db 427279 AACCGAACATCGCGGCGCAACATCGCGAC-----TTTAACTCGGATCGCA 427229
Qy 28 -----AlaProGlyIleLeuTyrValArgThrAspIleGly-----GlyMet 41
Db 427228 AACACCGGTCCGGGCTAACCGCGCTGTCAACAACATCGGTATCGGCAACACCGGCAAC 427169
Qy 42 TyrArgTrpAspAlaAlaAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60
Db 427168 TACAACATCGGTGTGCGCAACACCGGTAACTAACAC-----ATCGGCTTC 427124
Qy 61 AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp-----74
Db 427123 GGCACACCGCGCAACACATCGGATCGGCTGTCCGCGCAACACGATCGGTTTC 427064
Qy 75 ---ProIleAsnThrAsnLysValTrpAlaValGlyMetTyrThrAsnSerTrpAsp 93
Db 427063 GGCCGCTGAACGCGCGC-----ATCGCAACATGGGCTGCTTC---AACCTGGCGAC 427013
Qy 94 ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro 113
Db 427012 AACAACTTTGCG-----427001
Qy 114 LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal 133
Db 427000 -----ATGCCAACCGCGGCAACTTCAACCGGCGCATTTGCC 426965
Qy 134 AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp 152
Db 426964 AACACCGCGCAACACATCGGCTGTTCACACCGCGCAACACACGTCGCGCATCTGG 426905
Qy 153 ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr 172
Db 426904 CTGACCGCGCAGCGCTTGTCTCGGCTTCAGCTCCCTGAACTCCGCGCGCGCAACACCGGT 426845
Qy 173 TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp 192
Db 426844 TTCTCAACTCCGCGCACCGCAACACCGCGC-----426815
Qy 193 ValAlaPheAspLysSerSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal 211
Db 426814 ---TTGTTCACTCCGCGCACCGCAACACCGGCTGTTCACCTCGGCGCACCGCAACGTC 426758
Qy 212 GlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln 231
Db 426757 GGCATCGGCAACATGGCGCACCGCGGCTTCGCGCTCGGCTTCCGCGCGCACCGGAGTG 426698
Qy 232 AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn 251
Db 426697 GGCATCGGCGCACCAACTCGGCGAGTTTC---AACATCGGCTGTGTTTAACTCGGCGCAC 426641
Qy 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAs 271
Db 271 -----
```

Db 426640 GGCAATGTGGCATCGGCAACTCGGGCACCGCAACGTCGGCATCGGCAACACCGGCACC 426581  
Qy 271 pValTrpLysPheSerValThrSerGlyThr-----Tr 282  
Db 426580 GGCAACACCGGCATCGGAAACAGCGCACTACAAACCGCGCTTCTCAACCGCGGCCTG 426521  
Qy 282 p-----ThrArgLysProValProSerThrAspThrAlaAsnAs 296  
Db 426520 GTCAACACCGGCATCGGCAACCGGGCAACCAACACCGCGCTTCAACATCGGCACC 426461  
Qy 296 pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl 316  
Db 426460 TTCACACCGGCATCGGCAACCGGGCCACTACAAACCGCGCTTCTCAACACCGGTAGC 426401  
Qy 316 aThrGlnIleSerTrp-----TrpProAspThrIleIle-----PheArgSerThr----- 331  
Db 426400 TACAACACCGGCATCGGCAACCGCGGAGACTACGCGACCGCGCGCTTCTCATCCCGGACG 426341  
Qy 332 -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer-- 345  
Db 426340 ATGAACACCGCTTGTCTGTGGCGCGCCGACGCGGAGCGCTTGTGGCGGCCAATACACC 426281  
Qy 346 -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGl 360  
Db 426280 ATCACCATCGACGACCTGCGCGCTTCTCAATGTGCACATCCCGGTCAACATCCCATC 426221  
Qy 360 uProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTr 380  
Db 426220 ACCG---CGGACATCACCAATGTCTCCATCCCGCCATTACGTTCGCCAGA-ATC----- 426170  
Qy 380 pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh 399  
Db 426169 ---GAGCGACGCGAAGCGTCGACATAGGCATCTCAGTGGCACCGCTTGGCCCGCT 426114  
Qy 399 rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrIleTrpAspSerGlyGlnI 418  
Db 426113 CGTCCGATCATCCCTGTCATGCGGGGAGCGCTGCGCCCGCTGGGACACCCATCGAAT 426054  
Qy 418 eHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerPr 438  
Db 426053 TGACTTCGGCCCC-----TCGCGCGGATCAACCTCAACATCGGCA 426012  
Qy 438 oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAs 458  
Db 426011 GCCGACGCGCTCCACCGTGATCAACATCGTGGCGGCGCCGCG----- 425969  
Qy 458 pValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVa 478  
Db 425968 -----GCCGCGCCGATC-----AGCAT 425952  
Qy 478 lAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSe 498  
Db 425951 TCCGATCATCGACTTCGGCGCAGCG-----CCCGGCTTCTCAACCGCCACAC 425904  
Qy 498 rGlnProAsnAspArgHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGl 518  
Db 425903 CGGCCCG-----TCGTCGGCTTCTCAACTGGGTGGTGG 425868  
Qy 518 ySerGluProGlyGlyValThrThrGlyGlyThr-----ValAlaAl 532  
Db 425867 CAGCGCATCGGCTTGTGTAATCTCGGCAACAACTCGGGCCTCTACAACTTCGCCACTAG 425808  
Qy 532 aSerAlaAspGlySerArgPheVal-----TrpAlaProGl 544  
Db 425807 CAGCATGGGAATTCGGGCTTCCAAAATACTATGGGTGCTGCAGTTCGGGTGGCG----- 425753  
Qy 544 yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr 560  
Db 425752 -----AATTGGGCAACAGCATCTCGGCATCTA 425724  
Qy 560 pAlaAspGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy 580  
Db 425723 CAACACCGGCTTGGGAGCACCGGCAATGTC-----TCGGGCTTGTCTCAACATCGG 425673

Qy 580 sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVa 598  
Db 425672 CACCAACCTGGTGGTGGTTCAGAACCGGCG----- 425639  
Qy 598 lThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPhe- 617  
Db 425638 -ACCGAGACGACTTCAGCGTGGCTTGGCCAACTTCGGGTCT---GGAACTCGGGTAG 425583  
Qy 618 -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA 630  
Db 425582 CGCAACATCGGCAACTACAACTGGGACGCCCAACATCGGCGTCTACAACTGGGCAG 425523  
Qy 630 la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerA 647  
Db 425522 CGCCAACTCGGCGACTTCAACTGGGACGCGCAACATCGGCGACTTCAACCTGGGCAG 425463  
Qy 647 laIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerS 667  
Db 425462 CGCCAACTCGGCGACTTCAACTGGGACGCGCAACATCGGCGTCC-ACATCGGCTCGGCAACGTCGTCGGGG----- 425414  
Qy 667 eTyrProAlaValPheValValGlyThrIleGly---GlyValThrGlyValaTyrArgS 686  
Db 425413 -----CTGACGGGCGCATCGGCAACATCGGCTTCGGCAACCGGAAACGGAACA 425362  
Qy 686 eAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnT 706  
Db 425361 TCGGCATCGCAATACCGGCACCGCAACATCGGCTTCGGCAACCGGAAACGGAACA 425302  
Qy 706 rpGlyGlnAlaIleThrGlyAsp---HisAlaAsnLeuArgValTyrIle-GlyThrA 725  
Db 425301 TCGGCATCGGCTACCGCGCACCATGACCGGCTTCGGCGCTCGGAACCTCGGCACCG 425242  
Qy 725 snGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 425241 GCAACATCGGCTATTCAACTCGGCACCGGCAACATCGGCTTCGGCAACTCGGC 425186

## RESULT 7

US-09-902-540-1027/c  
; Sequence 1027, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1027  
; LENGTH: 10317  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-1027

## Alignment Scores:

|                        |          |               |       |
|------------------------|----------|---------------|-------|
| Pred. No.:             | 2,03e-05 | Length:       | 10317 |
| Score:                 | 184.00   | Matches:      | 155   |
| Percent Similarity:    | 29.4%    | Conservative: | 70    |
| Best Local Similarity: | 20.3%    | Mismatches:   | 228   |
| Query Match:           | 4.6%     | Indels:       | 312   |
| DB:                    | 3        | Gaps:         | 40    |

US-09-917-376-3 (1-740) x US-09-902-540-1027 (1-10317)

Qy 44 TrpAspAlaAlaAsnGlyArgTrpIlePro----- 53  
Db 4722 TGGGATCGCAGCGGAGGCTTCAATCATGACAGCGGACGGGCTCGGTGCAAAACGAG 4663





[illegible]

```

Qy 344 ThrSer-----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
Db 3743543 CGATCCCGGGCTTCACCATTCGCGCTCGCGCTG-ACGATCGACATAGACGGC 3743485
Qy 360 Glu----- 360
Db 3743484 CAGATTGACGGCTTCAGACACCCCGCATCAGCATCGACCGCATCCCGTGAACCTCGGC 3743425
Qy 361 -----ProTrrLeuThrPheGlyValGlnProAsnProProValPro 374
Db 3743424 GCCAGCGTCACTGTCGCGCTCTCTCATCAACCGCGTT-----AATATCCCGCGC-- 3743374
Qy 375 SerProLysLeuGlyTrrMetAspGluAlaMetAlaIleAspPropheAsnSer---Asp 393
Db 3743373 ACCCGCGGCTTGGCAACACACACCGCTCCGTCGCGGTTTCTTCAACTCCGCGCAC 3743314
Qy 394 ArgMetLeu-----TyrGlyThrGlyAlaThrLeuTyr-----Ala 405
Db 3743313 GGTGGGTGTCGGGCTTCGGGAATTCGTCGCGGCGAGCTCGGGTTGTGTGAACACGCGC 3743254
Qy 406 ThrAsnAspLeuThrTrrAspSerGly-----Gly 416
Db 3743253 CAGACCGAGGTGGTGGCGGGTTCGGTTTCGCCAATTCGTTTCGCTCGGTGCGGT 3743194
Qy 417 GlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIle 436
Db 3743193 GTGCTGAATTCGCTCGGTGTGTGCGGGTGTACACACCGCGCGG----- 3743146
Qy 437 SerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHis 456
Db 3743145 TTCCGCGCGGGACCCCGGGTGTCTCGGATCGGCAATGTTGTGTAG----- 3743095
Qy 457 AlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr 476
Db 3743094 -----CAGCTGTGCGGGTGTCTCGCGCGGACG 3743055
Qy 477 SerValAspTyrAlaGluLeuAsnProSerIleValAlaGlySerPheAspPro 496
Db 3743064 GCA-----CTCAACGAGCGCTCATCATCAATTCGCGGTGGCGCATGTG 3743020
Qy 497 SerSerGlnProAsnAspArgHisValAlaPhe-SerThrAspGlyGlyLysAsnTrpHis 516
Db 3743019 GGCAGCGTA-----AAGCTCGGTTTCGGCAACGTCGCGGACTTCAACCTCGGT 3742972
Qy 516 eGln-----GlySerGluProGlyGlyValThrGln 527
Db 3742971 CGCGCAATATCGCGACTTGAACGTGGTTCGGCAATGTCGCGCGGCAACGTGGG 3742912
Qy 527 YGlyThrValAlaAla-----SerAlaAspGlySerArgPheValTrpAlaProGlyAs 545
Db 3742911 TTCGGCAATATCGCGATGCAACTTCGGGTTCGGCAATCGCGGTCTGCGCGGCGCTG 3742852
Qy 545 pProGlyGlnProValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSer-----G1 564
Db 3742851 GCCCGGG-----TGGCAACATCGCGGTG 3742828
Qy 564 nGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAl 584
Db 3742827 GGCAATCGCGCGACCGCA-----AC 3742807
Qy 584 aLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAl 604
Db 3742806 GTCGGCTTCGGCAACATGGGTGTGGCAACATCGGGTTCGGTAAACCGCGCACCAACAC 3742747
Qy 604 aAlaGlyLeu-----ProSerSe 610
Db 3742746 CTCGGATTCGGGTGACCGGGGCAACACGATCGGATCGCGCTTGAACCTCGGT-GC 3742688
Qy 610 rGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyGlyAspLeuTrpLeuAl 630
Db 3742687 CGGCAACATCGGG-----TTGTTCAACTCCGCAACCGCAACGTCGGG-----TTGTTCAACT 3742634
Qy 630 aAlaSerSer-----GlyLeuTyrHisSer-----ThrAsnGlyGlySerSe 644

```

```

Db 3742633 CGGGACCGGGAACTTCGGGTGTTTCAACTCGCGCAGCTTCAACACCGGCATCGGCAATGG 3742574
Qy 644 rTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaPr 664
Db 3742573 CGGAACCGGGCAGTACTGGCTTTTCAATGCCGGTAAATTTCAATACGGGTGTGGCAACCC 3742514
Qy 664 oGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyValTyr 684
Db 3742513 TGGGTTCGTACAAACCGGCAGCTTCAATGTGGTGACACCAAC-----ACCGGTGGTTT 3742460
Qy 684 rArgSerAspAspCysGlyThrTrpValLeuIleAsnAspAspGlnHisGlnTyrG1 704
Db 3742459 CAACCCCGGCGAGCATCAACACCGGTGTTTCAACACCGGCACAGC-CAACACCGCGCTCG 3742401
Qy 704 yAsnTrpGlnAlaIleThrGlyAspHisAlaAsnLeuArgAsgValTyrIleGly 723
Db 3742400 CCAATTCGGGCA-----TGTCGACACCGCGCGCTCATGTGCGG 3742361

RESULT 9
US-09-328-1377
; Sequence 1377, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1377
; LENGTH: 11679
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1377

Alignment Scores:
Pred. No.: 4,11e-05 Length: 11679
Score: 181.50 Matches: 204
Percent Similarity: 31.0% Conservative: 90
Best Local Similarity: 21.5% Mismatches: 344
Query Match: 4.5% Indels: 309
DB: 3 Gaps: 54

US-09-917-376-3 (1-740) x US-09-328-352-1377 (1-11679)
Qy 7 ThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGly----- 21
Db 1627 ACGTGGACA---GTAAGTGTCCGCGGTAGTGTGTTGGTTGCTGATGCAGATAAAACGATT 1683
Qy 22 -----IleValPheAsnGluGlyAlaProGlyIleLeuTyrValArg---ThrAsp 37
Db 1684 GATGCTAAAGTAACGTTTACAGATGTCAGAGCTAATAGCAGCACTGTTAAACGATACGCAA 1743
Qy 38 IleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleLeuLeuAspTrp 57
Db 1744 ATT-----TATACATTAGACACAGCT---GCTCTCGCAGCGCCAGTAAATGACCCA 1791
Qy 58 ValGlyTrpAsnAsnTrpGlyTyrAsnGlyVal-----ValSerIleAlaAlaAspPro 75
Db 1792 GTT-----AACGGGACAGACCAATACAGGTACAGCAACACT 1830
Qy 76 IleAsnThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsn----- 90
Db 1831 GGTTCACAGTAACAGTA-----ACCTATCTCTAATGTTGTCACACCAACA 1875
Qy 91 -----SerTrpAsp---ProAsnAspGlyAlaIleLeuArgSer 102
Db 1876 GTTGTACAGACCGGCGAGTGTGTGTCAGTACCAACCCCTGGC-----CTTAATGAT 1929
Qy 103 SerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsn--- 121

```

Db 1930 GGTGACGAAGTTGAGGCAATTGTTACAGATCCA-----GCAGGCAACCA 1974  
Qy 122 ---MetProGlyArgGlyMetGlyGluArgLeuAlaValAlaAspProAsn----- 137  
Db 1975 TCTTTGCGAGGTACAGCTACTGTTGAT-----GCAGTTGGTCCAAATACCGATGGTGT 2028  
Qy 138 -----AspAsnIleLeuTyrPheGlyAlaProSerGly 148  
Db 2029 AACTTTACGGTTGATTGATTCAGTAAACAGCTCAGCAATGATTAATGATCAAGACGGTCAGGC 2088  
Qy 149 Lys-----GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
Db 2089 AACGTTACTGTTACTGTTGATTAAGAAACGTTCCGGCAGATGCGCAATACAGCTGTC 2148  
Qy 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
Db 2149 ACTGTTGTAATATGCGCAGCGTATCTGCAACTGTAGATAGCAGCAGGC----- 2202  
Qy 185 SerAspIleGlnGlyValTrpValAlaPheAspLysSerSerSerLeuGlyGln 204  
Db 2203 -----ACATGGACAGTAAGCGTACCAGGTAGTGACTTAACTCGGAT 2244  
Qy 205 AlaSerLysThrIle----- 209  
Db 2245 GCAGTAAGCAGATTGATGCTAAAGTAACGTTTACAGATCGGCGAGGTAATAGCAGCAGT 2304  
Qy 209 ----- 209  
Db 2305 GTTAAACGATACACAAATATACATACATCGATACCCTGACCTGATGCACCAAGTAATTAAAC 2364  
Qy 210 -----PheValGlyValAlaAspProAsnAsnProValPheTrp 222  
Db 2365 CCGGTTAAACGGGACAGACCCGATTACAGGTACGGCAGAGCCTGGTTCAACAGTAACCTGTG 2424  
Qy 223 SerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro---ThrGlyPheIle 241  
Db 2425 ACTTATCCAGATGCGCATACACACAGCTGTTGCGAGCCGGATGGCAGTGGACAGTA 2484  
Qy 242 ProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThr 261  
Db 2485 CCAACCCAGGTTTAATGATGGC---GATAAAGTTACAGCAATGCTACAGATCCAGCA 2541  
Qy 262 GlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThr 281  
Db 2542 GGCAACCA-----TCATTACCAGGTACA 2565  
Qy 282 TrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
Db 2566 GCTACTGTTGATGCACTGGTTCCTCAAAATACCGATGGTGTAAAT-----TTCACGGTTGAT 2619  
Qy 302 GlyLeuThrIleAspArg----- 307  
Db 2620 TCAGTAACAGCTGACAATGATTAATGATCATCAGAAGCATCAGGCAACGTTACTGTTTACA 2679  
Qy 308 -----GlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 2680 GGTGATTGAAACGTTCCGGCAGATCCAGCAATACAGT---GTCACCTGTTGTGATC 2736  
Qy 320 SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg 339  
Db 2737 AATGCCAGACGTTACTGCAACTGTAGATAGCAGCAGGC-----ACATGGACA--- 2787  
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359  
Db 2788 -----GTAAGCGTACCAGGTAGTGACTTGACTGCGGATGTCAGATAAGACGATT 2835  
Qy 360 GluProTrpLeuThrPhe----- 365  
Db 2836 GATGCCAAAGTAACGTTTACAGATGCGCAGCGTAACACAGCAGGTGTTAAACGATACACAC 2895  
Qy 366 -----GlyValGlnProAsnProProValProSerProLysLeuGlyTyr 380  
Db 2896 ACATATACAGTTGATACGGTTGCACCAAAATGCACCGGTG----- 2934

Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAsp----- 393  
Db 2935 CTTGATCCGATCAATGCAACAGACCCAGCTGAGCGGTGAGCAGAGCCCTGGTTCAACAGTG 2994  
Qy 394 ArgMetLeuTyr-----GlyThrGlyAlaThrLeuTyrAlaThrAsnAsp----- 408  
Db 2995 ACTGTGACTTATCTCTGATGGCACTGTCACACAGTGTGTAGCAGGACCGGATGGTAGCTGG 3054  
Qy 409 -----LeuThrLysTrpAsp 413  
Db 3055 TCAGTACCAAAACCCAGGTAACCTGGTGGATGGCGATACAGTACTGCAACAGCAACTGAC 3114  
Qy 414 SerGlyGlyGlnIleHisIle-----AlaPro 422  
Db 3115 CCTGCGAGCAACACTTCTATTCGAGGTACAGGCACAGTTCCTCAGCAGACATCACAGCACCT 3174  
Qy 423 MetValLysGlyLeuGluGlnThrAlaValAsnAsp-----Leu 435  
Db 3175 GTGGTT---GCCTGGATGACGCTGTGCAAGATGACAGCACACACAGCACTTACCGGTACA 3231  
Qy 436 IleSerProSerGlyAlaProLeuIleSerAla----- 447  
Db 3232 GTGAAGATCCGACAGCCACTGTAGTTGCAATGTGATGGCTGTGACTATCCGGCAGTG 3291  
Qy 448 ---LeuGlyAspLeuGlyGlyPheThrHisAlaAsp-----ValThrAla 461  
Db 3292 AACAAATGGTGAC---GGCACCTGGACGCTTGCAGACAATACACTTCTCGGTTAACTGAT 3348  
Qy 462 ValProSerThrIlePheThrSerProValPheThrThrGly----- 475  
Db 3349 GGTGCACACACACTTACCTGCTGCTCAACAGATGCGAGCGCAATGCGAGGTACAGATACA 3408  
Qy 476 -----ThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493  
Db 3409 GCAGTGGTGACGATTGATACCAAGCAAAATGCAACCGGTACTGTATCCGATCAATGCG 3468  
Qy 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513  
Db 3469 ACTGACCCA-----GTGAGCGGTACAGCAGAGGTGGTTCA 3504  
Qy 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533  
Db 3505 ACC---GTGACTGTGACTTATCCTGATGCGACCACT---GCAACAGTGGTAGCAGGC 3555  
Qy 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProVal----- 550  
Db 3556 ACAGATGGTAGC-----TGGTCAGTACCAACCCAGGTAACTCTGGTAGATGGTAT 3606  
Qy 551 ---ValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569  
Db 3607 ACAGTAACCTGCAACA-----GCAACTGACCTGCGAGCAATACATCATTTGCCAGGTACA 3660  
Qy 570 AlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589  
Db 3661 GGCACAGTCTCTGCAGACATCAGACACCT-----GTGGTTGGCTGGATGCGTGTG 3714  
Qy 590 PheTyrArgSerThrAspGly-----GlyValThrPheGlnProValAlaAlaGlyLeu 607  
Db 3715 ACCAATGACAGCACACACAGCACTTACGGGTACAGTGAACCGATCCGACAGCACTGTAGTG 3774  
Qy 608 ProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeu 627  
Db 3775 GTGAAT-----GTAGATGGCAGTCACTATCCGGCAGTGAACAAATGCGCAGCGC---ACC 3825  
Qy 628 TrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla 647  
Db 3826 TGGACGCTTGCAGACATACACTTCCAGTGTGGCAGACGCTCCACACAC---ATTACC 3882  
Qy 648 IleThrGlyValSerSerAlaValAsnValGly-----PheGly 660  
Db 3883 GTGACTGCAACAGATGCGACAGGCAATGCGAGGTACAGATACAGCAGTGGTGACGATTGAT 3942

```

Qy 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyVal 680
Db 3943 ACCACGACCAATGCA-----CGGTACTTGCATCAATGCGACTGCCAGTG 3996
Qy 681 ThrGlyAlaTyrArgSerAspCys-----GlyThrThrTrp 693
Db 3997 AGCGGTACGACGAGCTGTTCAACGGTGACTGTGACTTATCCTGATGCCACTGCA 4056
Qy 694 ValLeuIle-----AsnAspAspGlnHisGlnTyrGlyAlaIleThr 711
Db 4057 ACAGTGGTAGCGACACAGATGGTGTGTGCTAGTACCAACCCAGGTAACCTGTAGT 4116
Qy 712 GlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsn-----GlyArg 727
Db 4117 GGTGATACAGTAAGTGCACAGCAACTGACCTGCAGGCAATACATTCATGCCAGGTACA 4176
Qy 728 GlyIleValTyrGlyAspIle 734
Db 4177 GGCACAGTCTCTGCAGACATC 4197

RESULT 10
US-09-949-016-11814
; Sequence 11814, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11814
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-11814

Alignment Scores:
Pred. No.: 0.000436 Length: 29927
Score: 177.00 Matches: 189
Percent Similarity: 28.8% Conservative: 56
Best Local Similarity: 22.2% Mismatches: 292
Query Match: 4.4% Indels: 314
DB: 3 Gaps: 48

US-09-917-376-3 (1-740) x US-09-949-016-11814 (1-29927)

Qy 14 GlyGlyGlyGlyPheValAspGly-----IleValPheAsnGluGly 27
Db 19367 GGGGGCTCTGGCAGGCGGACGGTCCCTTCTTCCACCGCTTTTATTCGAAGGGGAC 19426
Qy 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41
Db 19427 AGGCTGGGAGTGTATTGGCGCGGTGTGGCTGAGGTGCAGGACTT-GGGGGG--- 19488
Qy 42 TyrArgTrpAspAlaAlaAsn----- 48
Db 19483 TGGCGGTGGGAGCGCGGAAGGTATAACGTATAATCATAGTAACAACTCAGAAATG 19544
Qy 49 -----GlyArgTrpIleProLeu-LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAs 66

```

```
QY 369 oAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPr 389
Db 20393 ---CCTGGTCTCCGGGAGGCTTGGGAGCTTGGTCTCCGGGAGGCTTGGACACC 20448
QY 389 o-PheAsnSerArgMetLeuTyrglyThrGlyAlaThrLeuTyralaThrAsnAspL 409
Db 20449 TGGTCTCCAGAGAGGCTT-----GGGACCTGGTGTCTCCGGAGGCTTGGAGGCTGGGGACC 20502
QY 409 euThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429
Db 20503 TGGTCAACCGGAGAGCCTTGGG----- 20525
QY 429 luThrAlaValAsnLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447
Db 20526 -----GACCTGGTGTCTGGGAGAGCCTTGGGGACCT----- 20558
QY 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467
Db 20559 -----GGTACCTTGGAGAGCTTGGGAGCCTGGTGTCTCCGGAGTGCCTTGGGACCT- 20612
QY 467 heThrSerProValPheThrThrGlyThrSerValAspTyralaGluLeuAsnProSerI 487
Db 20613 --AGTGACCGGGAGAGCCTTGGGACCTGTGTCTCCGGGAGAGCCTTGGGACCTGTG 20670
QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507
Db 20671 TCCTG-----GGAGAGCCTTGGGGATCTGGTGTCTCTG- 20702
QY 507 heSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG 527
Db 20703 -----GGGAGAGCTGGGGAGCCTGTGTCTCCGGAGAGAGCCTTGGGGACCTGTGACC 20757
QY 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP 546
Db 20758 CGGGAGGCTTGGACACCTGGTGTCTCCGGGAGAG-----GCTTGGGGACCTGTGTGACC 20811
QY 546 roGlyGln-----ProValValTyralaValGlyPheGlyAsnSerTrpAlaAlaAs 563
Db 20812 CGGGAGAGCCTTGGGAGCCTGTGTCTCCGGGAGAGGCTGGGG---ACCTGGTGTCTCG 20868
QY 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgValA 578
Db 20869 GGAGAGAGCCTTGGGGACCTGTGTACCCGGGAGAGGCTTGG----- 20909
QY 578 snProLysThrPheTyralaLeuSerAsnGlyThrPheTyArgSerThrAspGly---- 596
Db 20910 -----ACACCTGGTGTGCC 20922
QY 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615
Db 20923 CGGGAGGCTTGGAGAGCCTGGTGTCCGGGAGAGCCTTGGGGACAGTACCTTGGAG 20982
QY 615 alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyL 635
Db 20983 AGGCTTGGGGA-----CCTGGTGTCTGGAGAGCCTTGGGACCTGGTGTCTCGGA- 21035
QY 635 euTyrrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlav 655
Db 21036 -----GAGGTTACGGGGGCTGTGTGGGGAGAGACGTTGTGAGGCAAAAGTCCC 21084
QY 655 alAsn----- 656
Db 21085 TGAATCCCTCGAAGAGAGCGATCGGGAGCTCCCTCCCTGAGGGCGGTCATTTGGGACC 21144
QY 657 -----ValGlyPheGlyLysSerAlaProGlyLysSerSerTyrrProAlav 671
Db 21145 CCCCTCCATCGCTTTCAGAGGAGCTTTCGGATTCCCTTGGCCCGGCT---CCCGGG 21201
QY 671 alPheValValGlyThrIleGlyValThrGlyAlaTyrrArgSerAspCysGlyT 691
Db 21202 ATGCATCCAGTGGCAGCCCAATTCGTGGCCAGCGGGAGGAGGAAAGCGGGTGTGGG 21261
```

```
QY 691 hrThrTrpValLeuIleAsnAspGlnHisGlnTyrrGlyAsnTrpGlyGlnAlaIleT 711
Db 21262 TGGTC-----TCACGG 21273
QY 711 hrGlyAspHisAlaAsnLeuArgValTyrrIleGlyThrAsnGlyArgGlyIleValT 731
Db 21274 CTGGAGAAGGGGCGACGCTCCCTAGG-----GGAGAGGAGGACGCTTGGGGGTTTC 21324
QY 731 yrGlyAspIleGlyGlyAla 737
Db 21325 CGGGGGCGGGCGGCGGAGCA 21344

RESULT 11
US-09-949-016-17474
; Sequence 17474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17474
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17474

Alignment Scores:
Pred. No.: 0.000436 Length: 29927
Score: 177.00 Matches: 189
Percent Similarity: 28.8% Conservative: 56
Best Local Similarity: 22.2% Mismatches: 292
Query Match: 4.4% Indels: 314
DB: 3 Gaps: 48

US-09-917-376-3 (1-740) x US-09-949-016-17474 (1-29927)
QY 14 GlyGlyGlyGlyPheValaspGly-----IleValPheAsnGluGly 27
Db 19367 GGGGGCTCTGGCAGGCGGCGGCGGCTCCCTCTTCCACCGTTTTTATTTCCAGGGGAC 19426
QY 28 AlaProGlyIleLeuTyrrValargThr-----AspIleGlyGlyMet 41
Db 19427 AGGCTGGGGAITGTATTGTGGCGCGTGTGGCTGAGGTGCGAGGACTT-GGGGG- 19482
QY 42 TyrrArgTrpAspAlaAlaAsn----- 48
Db 19483 TGGCGGTGGGAGCGCGGAGGTATAAAGCTATAATCATATAAGTAACAACACTCAGAAATG 19542
QY 49 -----GlyArgTrpIleProLeu-LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrrAs 66
Db 19543 GACCCCGAGCGCTGGTGGCGGCTAGCTCTCCAGCTCTCCCTGGCCCGGCGGAGGAGA 19602
QY 66 nGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValG 86
Db 19603 GGGGTCCGCATCCCTCCCGGCTTCTCTCTC----- 19633
QY 86 yMetTyrrThrAsnSerTrpAspProAsnAspGlyAla-----IleLeuArgSerSe 103
Db 86 yMetTyrrThrAsnSerTrpAspProAsnAspGlyAla-----IleLeuArgSerSe 103
```

Db 19634 -----CTGGGTACCTGGCTTGGAGTGGGGAGAACGAGCCTACTCTTCTGTACCGTCTTT 19686  
Qy 103 rAspGlnGlyAlaThr-----TrpGlnI1 111  
Db 19687 TGCCGACGGCGGACCCAGTGAATTAGCCGTGGAGCCCGCAGGCTGCCTGGCTTTG 19746  
Qy 111 eThrProLeuProPheIysLeuGlyGlyAsnMetProGlyArgGlyMetGly-----128  
Db 19747 CGCACCGGAGCTCTGG-----GGACCTGTGTCTCCCGGAGAAACCTGGGGACCTGGTA 19800  
Qy 129 -----GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr-----142  
Db 19801 TCCCGGGAGAGCTTGGGACCTGGTGTCTCCGGAGAGGCTGGGTACTGTCTCTCT 19860  
Qy 143 -----PheGlyAlaProSerGlyIysGlyLeu-----151  
Db 19861 GGAAGAGGCTTGACACCTGGTGTCTCTGGAGGCTTGGACCTGTGTCTCTGGGAGA 19920  
Qy 152 -----TrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProaspVa 170  
Db 19921 GGCTTGGAGATCTGTGTCTCTGGGAGAGGCTTGGGGA-----19957  
Qy 170 lGlyThrTyIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190  
Db 19958 -----CCTGGTGTCTCTGGAGA 19974  
Qy 190 lValTrpValAlaPheAspIysSerSerSerLeuGlyGlnAlaSerIysThrIlePh 210  
Db 19975 GGCTTGG-----GGACCTGGTGACCTTGGAGAGGCTTGGAGACCTGGTGT 20019  
Qy 210 eValGlyAlaAlaAspProAsnAsnPro-ValPheTrpSerArg-----224  
Db 20020 TCTGGGAGAGGCT-----TGGGAGACCTGTGTCTGGGAGAGGCTTGGGACCTGGTGT 20073  
Qy 225 -----AspGlyGlyAlaThrTrp-----GlnAlaValProG 235  
Db 20074 CTCTGGAAGAGGCTTGACACCTGGTGACCCGGAGGCTTGGGGATCTGGTGTCTCCGG 20133  
Qy 235 lAlaProThrGlyPheIleProHisIysGlyValPheAspProValAsnHisValLeut 255  
Db 20134 GAGAGCCTTGGGA-----CCTGGTGTCTCTG-----20159  
Qy 255 yrIleAlaThrSerAsnThrGlyGlyProTyTrpAspGlySerSerGlyAspValTrpIysP 275  
Db 20160 -----CGAGAGGCTTGGGACCT-----GGTGACCTTGGAGAGGCTTGGGAC 20202  
Qy 275 heSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaA 295  
Db 20203 CTGGTGTCTCTGAGAGCCTTGGGGA-----TCTGGTGTCTCCAGGAGAGGCTTGGGAC 20256  
Qy 295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315  
Db 20257 CT-----GGTGTCTCTGGAAGGCTTGGAC-----20282  
Qy 315 alAlaThrGlnIleSerTrpTrpPro-----AspThrIleIlePheArgs 330  
Db 20283 -----ACCTGGTGTCTCTGGGAGAGGCTTGGGACCTGGTG-----20318  
Qy 330 erThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp-TrpThrSerTyrProAsnArg 349  
Db 20319 -----TCTGGGAGAGGCTTGGGACCTGGTGT-----CCTGGGAGA 20355  
Qy 350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPr 369  
Db 20356 GGCTTGGAGA-----TCTGGTGTCTGGGAGAGGCTTGGGGA-----20392  
Qy 369 oAsnProValProSerProIysLeuGlyTyrMetAspGluAlaMetAlaIleAspPr 389  
Db 20393 -----CCTGGTGTCTCCGGGAGAGGCTTGGGACCTTGGTCTCCGGGAGAGGCTTGGACACC 20448  
Qy 389 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409  
Db 20449 TGTGTCTCCAGGAGAGGCTT-----GGGAGACCTGGTGTCTTGGAGAGGCTTGGGAGC 20502

Qy 409 euThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValIysGlyLeuGluG 429  
Db 20503 TGGTGACCCCGGAGAGCCTTGGG-----20525  
Qy 429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447  
Db 20526 -----GACCTGGTGTCTCTGGGAGAGCCTTGGGACCT-----20558  
Qy 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467  
Db 20559 -----GGTGACCTTGGAGAGCCTTGGGACCTGTGTCTCGGAGTGTCTTGGGACCT- 20612  
Qy 467 heThrSerProValPheThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487  
Db 20613 ---AGTGACCCCGGAGAGGCTTGGGACCTGTGTCTCCGGGAGAGGCTTGGGACCTGGTG 20670  
Qy 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaP 507  
Db 20671 TCCTG-----GGAGAGCCTTGGGGATCTGTGTCTCTG- 20702  
Qy 507 heSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG 527  
Db 20703 -----GGGAGAGGCTTGGGACCTGTGTCTCGGAGAGAGCCTTGGGACCTGGTGACC 20757  
Qy 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP 546  
Db 20758 CGGAGAGGCTTGGACACCTGTGTCTCCGGGAGAG-----CCTTGGGACCTGGTGACC 20811  
Qy 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlas 563  
Db 20812 CGGAGAGCCTTGGGAGCCTGTGTCTCTCGGAGAGGCTTGGGG---ACCTGTGTCTCG 20868  
Qy 563 erGln-----glyValProAlaAsnAlaGlnIleArgSerAspAsgValA 578  
Db 20869 GGAGAGAGCCTTGGGAGCCTGTGTCTCGGAGAGAGCCTTGG-----20909  
Qy 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly---- 596  
Db 20910 -----ACACCTGGTGTC 20922  
Qy 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615  
Db 20923 CGGAGAGGCTTGGGAGCCTGTGTCTCCGGGAGAGCCTTGGGACAGGTGACCTTGGAG 20982  
Qy 615 alMetPheHisAlaValProGlyLysGlnGlyAspLeuTrpLeuAlaAlaSerSerGlyL 635  
Db 20983 AGGCTTGGGGA-----CCTGTGTCTCTGGAGAGGCTTGGGACCTGTGTCTCGGGA- 21035  
Qy 635 euTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaV 655  
Db 21036 -----GAGTTACGGGGCTGTGTGGGGAGAGACGTTGTGAGCCAAAGTCCC 21084  
Qy 655 alAsn-----656  
Db 21085 TGAATCCTCGAAGAGAGCGCATCGGAGCTCCCTCGAGGCGGTCCATTTGTGGACC 21144  
Qy 657 -----ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaV 671  
Db 21145 CCCCTCCATCGCTTTTGCAGGAGCTGTTCGGATTCCTCTGGCCCGCT---CCCGCGG 21201  
Qy 671 alPheValValGlyThrIleGlyValThrGlyAlaTyrArgSerAspAspCysGlyT 691  
Db 21202 ATGCATCCAGTGGCAGCCCAATTCCTGGCCAGGGGAGAGGAAAGCGCGGTGTGGG 21261  
Qy 691 hrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleT 711  
Db 21262 TGCTC-----TCCACGG 21273  
Qy 711 hrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValT 731  
Db 21274 CTGGAGAAGGGCGACGCTCCCTAGG-----GGAGAAGAGGACGCTTGGGGGTTTC 21324



```

QY 731 yrGlyAspIleGlyGlyAla 737
Db 21325 CGGGGGCGGGGGCGGAGCA 21344

RESULT 12
US-09-949-016-17475
; Sequence 17475, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17475
; LENGTH: 29927
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(29927)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17475

Alignment Scores:
Pred. No.: 0.000436 Length: 29927
Score: 177.00 Matches: 189
Percent Similarity: 28.8% Conservative: 56
Best Local Similarity: 22.2% Mismatches: 292
Query Match: 4.4% Indels: 314
DB: 3 Gaps: 48

US-09-917-376-3 (1-740) x US-09-949-016-17475 (1-29927)
QY 14 GlyGlyGlyGlyPheValAspGly-----IleValPheAsnGluGly 27
Db 19367 GGGGCTCTGGCAGGGCGGCGGTGGCTCCCTCTTCACCGTTTATTCAGGGGAC 19426
QY 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41
Db 19427 AGGCTGGGGATTGTATTGGGCGGTGTTGGCTGAGGGTGCGAGGACTT-GGGGG-- 19482
QY 42 TyrArgTTrpAspAlaAlaAsn----- 48
Db 19483 TGGCGGTGGGAGCGCGGAGGTATAAAGTATAATCATAGTAACAACATCGAAATG 19542
QY 49 -----GlyArgTTrpIleProLeu-LeuAspTrpValGlyTTrpAsnAsnTTrpGlyTyrAs 66
Db 19543 GACCCGAGCGCTGTCCCGCTAGCTCTCCAGCTCTCCCTGCGCCAGCGCCGGAAGGAGA 19602
QY 66 nGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTTrpAlaAlaValGl 86
Db 19603 GGGGTCCGCATCCCTCCCGGGTTCCTCTC----- 19633
QY 86 yMetTyrThrAsnSerTTrpAspProAsnAspGlyAla-----IleLeuArgSerSe 103
Db 19634 -----CTGGGTACCTGGCTTGAAGTGGGGACGAGCCTACTTCTGTGACCTTTT 19686
QY 103 rAspGlnGlyAlaThr-----TTrpGlnI 111
Db 19687 TGCCGACGGCGGACCCAGTGAATTAGCCGCTTGGACCGCGAGCCTGCTGCTGTTG 19746
QY 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGly----- 128

```

```

Qy 447 laLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467
Db 20559 -----GGTGACCTTGGAGAGCTTGGGAGCCTGTGTCTCGGAGTGCCTTGGGAGCCT- 20612
Qy 467 heThrSerProValPheThrThrThrThrSerValAspTyrAlaGluLeuAsnProSerI 487
Db 20613 --AGTGACCGGGAGAGCTTGGGAGCCTGTGTCTCGGAGAGCTTGGGAGCCTGGT 20670
Qy 487 lelleValArgAlaGlySerPheAspProSerGlnProAsnAspArgHisValAlaP 507
Db 20671 TCCTG-----GGAGAGCCTTGGGATCTGGTCTCTG- 20702
Qy 507 heSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG 527
Db 20703 -----GGGAGAGCTTGGGAGCCTGTGTCTCGGAGAGCCTTGGGAGCCTGGTACC 20757
Qy 527 lyGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP 546
Db 20758 CGGAGAGCCTTGGACACCTGGTGTCTCGGAGAG-----GCTTGGGAGCCTGTGTACC 20811
Qy 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaS 563
Db 20812 CGGAGAGCCTTGGGAGCCTGTGTCTCGGAGAGCCTGGG-----ACCTGTGTCTCG 20868
Qy 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgValA 578
Db 20869 GGAGAGAGCCTTGGGAGCCTGTGTCTCGGAGAGCTTGG----- 20909
Qy 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly---- 596
Db 20910 -----ACACCTGTGTCTCC 20922
Qy 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615
Db 20923 CGGAGAGCCTTGGGAGCCTGTGTCTCGGAGAGCCTTGGGACCAAGGTGACCTTGGAG 20982
Qy 615 alMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyL 635
Db 20983 AGGCTTGGGA-----CCTGTGTATCTTGGAGAGCCTTGGGAGCCTGTGTCTCGGA- 21035
Qy 635 euTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlav 655
Db 21036 -----GAGTTACGGGGCTGTGTGGGAGAGAACCTTGTGAGCAAAAGTCCC 21084
Qy 655 alaSn----- 656
Db 21085 TGATCCCTCGAAGAGAGCGCATCGGAGCTCCCTCGAGGGCGTTCCATTTGTGGACC 21144
Qy 657 -----ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlav 671
Db 21145 CCCCTCCATGCGCTTTCAGGAGCTGTTCGATTCCTTCCCTGCGCCGCT---CCGCGG 21201
Qy 671 alPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyT 691
Db 21202 ATGCATCCAGTCGAGCGCAATCTCTGGGCCAGGAGGAGAAAGCGGTGTGGG 21261
Qy 691 hrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyAlaIleT 711
Db 21262 TGCTC-----TCCACGG 21273
Qy 711 hrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValT 731
Db 21274 CTGAGAGAGGGGAGCCTCCCTAGG-----GGAGAAGAGGACGCTTGGGGGTTC 21324
Qy 731 yrGlyAspIleGlyGlyAla 737
Db 21325 CGGGGGCGGGGGGAGCA 21344

```

RESULT 13

US-09-949-016-12639

; Sequence 12639, Application US/09949016

; Patent No. 6812339

```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12639
; LENGTH: 37802
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(37802)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-12639

```

## Alignment Scores:

```

Pred. No.: 0.000623 Length: 37802
Score: 177.00 Matches: 189
Percent Similarity: 28.8% Conservative: 56
Best Local Similarity: 22.2% Mismatches: 292
Query Match: 4.4% Indels: 314
DB: 3 Gaps: 48

```

US-09-917-376-3 (1-740) x US-09-949-016-12639 (1-37802)

```

Qy 14 GlyGlyGlyGlyPheValAspGly-----IleValPheAsnGluGly 27
Db 19367 GGGGGCTCTGGCAGCGCGGACGCGTGGCTCCCTTCTTCCAGCTTTCACAGGGGAC 19426
Qy 28 AlaProGlyIleLeuTyrValArgThr-----AspIleGlyGlyMet 41
Db 19427 AGGCTGGGGATGTATTGTGGCGCGTGTGTGGCTGAGGTGCGAGGACTT-GGGGG- 19482
Qy 42 TyrArgTrpAspAlaAlaAsn----- 48
Db 19483 TGGCGTGGGGAGCGCGAGGTATAACTAATCAATAGTAACACTCAGAAATG 19542
Qy 49 -----GlyArgTrpIleProLeu-LeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAs 66
Db 19543 GACCCCGAGCGCTGTCGCGCTAGCTTCCAGCTCTCCCTGGCCCGCAGCGGAGAGA 19602
Qy 66 nGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValG 86
Db 19603 GGGGTCCGATCCCTCCGCGTCTCTCTC----- 19633
Qy 86 yMetTyrThrAsnSerTrpAspProAsnAspGlyAla-----IleLeuArgSerSe 103
Db 19634 -----CTGGTACCTGGCTTGGGTGGGGAAACGAGCCTACTTCTGTACCGCTTT 19686
Qy 103 rAspGlnGlyAlaThr-----TrpGlnI 111
Db 19687 TGGCGAGCGGGGAGCCAGTGAATAGCCGTGGAGCCCGAGGCTGCTGGCTTTG 19746
Qy 111 eThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGly----- 128
Db 19747 CGCACCGAGTCTTGG-----GGACCTGGGTCCCCGGGAAACACTTGGGACCTGGA 19800
Qy 129 -----GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyr----- 142
Db 19801 TCCCGGAGAGGCTTGGGACCTGGTGTCTCCGAGAGAGGCTTGGGTACCTGTTCTCT 19860
Qy 143 -----PheGlyAlaProSerGlyLysGlyLeu----- 151

```

19861 GGAAGAGGCTTGGACACCTGCTGCTTGGAGGGCTTTGGACCTGCTGCTCTGGGAGA 19920  
152 ----TTPArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnProAspVa 170  
19921 GCGTTGGAGATCTGTTGCTGGGAGAGGCTTGGGA----- 19957  
170 lGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyVa 190  
19958 ----- 19974  
190 lValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePh 210  
19975 GGCITGG-----GGACCTGGTGACCTTGGAGAGGCTTGGACACCTGGTGT 20019  
210 eValGlyValAlaAspProAsnAsnPro-ValPheTrpSerArg----- 224  
20020 TCTGGAGAGGCT-----TGGGACCTGGTGTCTTGGGAGAGGCTTGGGACCTGGTGT 20073  
225 -----AspGlyGlyAlaThrTrp---GlnAlaValProG 235  
20074 CTCTGGAAGAGGCTTGGACACCTGGTGACCGGGAGGCTTGGGATCTGGTGTCCCGG 20133  
235 lYAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeu 255  
20134 GAGAGCCTTGGGA-----CCTGTGTCTCTG----- 20159  
255 yRIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysP 275  
20160 -----GGAGAGGCTTGGGACCT-----GGTGACCTTGGAGAGGCTTGGGAC 20202  
275 heSerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaA 295  
20203 CTGGTGCTCCTGAGAGAGCCTTGGGA-----TCTGGTGCTCCAGAGAGGCTTGGGAC 20256  
295 snAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetV 315  
20257 CT-----GGTGCTCTCGAAGAGGCTTGGAC----- 20282  
315 alAlaThrGlnIleSerTrpTrpPro-----AspThrIleIlePheArgs 330  
20283 -----ACCTGTGTCTTGGGAGAGGCTTGGGACCTGGTG----- 20318  
330 erThrAspGlyGlyAlaThrTrpThrArgIleTrpAsp-TPThrSerTyrProAsnArg 349  
20319 -----TCTCGGAGAGGCTTGGGACCTGTGT-----CCTGGGAGA 20355  
350 SerLeu-ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnPr 369  
20356 GCGTTGGAGA-----TCTGTGTAGCCGGGAGAGGCTTGGGA----- 20392  
369 oAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPr 389  
20393 -----CCTGGTGCTCCGGGAGAGGCTTGGGACCTGGTGCTCCGGGAGAGGCTTGGACACC 20448  
389 o-PheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspL 409  
20449 TGGTGCTCCAGAGAGGCTT-----GGGACCTGGTGACCTTGGAGAGGCTTGGGAC 20502  
409 euThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluG 429  
20503 TGGTGACCCGGGAGAGGCTTGGG----- 20525  
429 luThrAlaValAsnAspLeuIleSer-----ProProSerGlyAlaProLeuIleSerA 447  
20526 -----GACCTGTGTCTTGGGAGAGGCTTGGGACCT----- 20558  
447 laLeuGlyAspLeuGlyPheThrHisAlaAspValThrAlaValProSerThrIleP 467  
20559 -----GGTGACCTTGGAGAGGCTTGGGACCTGGTGCTCGGAGAGTGCCTTGGGACCT- 20612  
467 heThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerI 487  
20613 --AGTGACCCGGGAGAGGCTTGGGACCTGTGTCTCCGGGAGAGGCTTGGGACCTGTGTG 20670

QY 487 leIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlap 507  
Db 20671 TCCTG-----GGAGAGCCTTGGGATCTGGTCTCTCTG- 20702  
QY 507 heSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrG 527  
Db 20703 -----GGAGAGGCTTGGGACCTGTGTCTCGGAGAGAGCCTTGGGACCTGGTGACC 20757  
QY 527 lYGlyThrVal---AlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspP 546  
Db 20758 CGGGAGAGCCTTGGACACCTGTGTCTCCGGGAGAG-----GCTTGGGACCTGGTGACC 20811  
QY 546 roGlyGln-----ProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlas 563  
Db 20812 CGGGAGAGCCTTGGGACCTGTGTCTCGGAGAGAGCTTGGGG-----ACCTGTGTCTCG 20868  
QY 563 erGln-----GlyValProAlaAsnAlaGlnIleArgSerAspArgVala 578  
Db 20869 GGAGAGAGCCTTGGGACCTGTGTGACCCGGGAGAGGCTTGG----- 20909  
QY 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGly--- 596  
Db 20910 -----ACACCTGGTGCTCC 20922  
QY 597 --GlyValThrPheGlnProValAlaAlaGlyLeuPro---SerSerGlyAlaValGlyV 615  
Db 20923 CGGGAGAGCCTTGGGACCTGTGTCTCCGGGAGAGCCTTGGGACAGGCTGACCTGGAG 20982  
QY 615 alMetPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGlyL 635  
Db 20983 AGGCTTGGGGA-----CCTGTGTATCTTGGAGAGGCTTGGGACCTGTGTCTCGGA- 21035  
QY 635 euTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlav 655  
Db 21036 -----GAGTTACGGGGCTGTGTGGGGAGAGACGTTGTGAGCAAAAGTCC 21084  
QY 655 alaSn----- 656  
Db 21085 TGAATCCCTGCGAAGAGAGCGCATCGGAGCTCCCTCGAGGCGCTTCCATTTGTGGACC 21144  
QY 657 -----ValGlyPheGlyLysSerAlaProGlySerSerTyrProAlav 671  
Db 21145 CCCCTCCCATCGCTTTCGAGGAGCTGTTCGATTCCTCCCTGCGCGGCT---CCGCGG 21201  
QY 671 alPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyT 691  
Db 21202 ATGCATCCAGTGGCAGCGCCCAATTCGCGCCAGGGGGAAGGAGGCGGCTGTGGGG 21261  
QY 691 hrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 711  
Db 21262 TGGTC-----TCCACGG 21273  
QY 711 hrGlyAspHisAlaAsnLeuArgArgValTyrIleGlyThrAsnGlyArgGlyIleValT 731  
Db 21274 CTGGAGAGGGGCGAGCGTCCCTAGG-----GGAGAAGAGGACCGTGGGGGTTTC 21324  
QY 731 yRGlyAspIleGlyAla 737  
Db 21325 CGGGGCGCGGGCGGAGCA 21344

## RESULT 14

US-09-902-540-2590  
; Sequence 2590, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540

| Seq ID | Seq Name  | Length | Score  | Percent Similarity | Best Local Similarity | Query Match | DB |
|--------|---|--------|--------|--------------------|-----------------------|-------------|----|
| 2872   | AATGCC  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 288    | ValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg  | 165    | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 2908   | ACGCCGGTG   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 308    | GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIle  | 165    | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 2932   | ACCACGCCGACGCGACGTTTCATCGCGCCGAGGTACAG                        | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 328    | PheArgSerThr  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 2989   | TTCCGGTCTACGGTCAGCGACGCGCATCGCCACGGTGAACG                     | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 345    | rTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3034   | GTACACGTCACGGTCCACACGCGA                                      | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 365    | eGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMe     | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3071   | TCGTCTCCGACACGCTCCGTCGCGCGGCACT-ATCACCCTGACGGCCTC             | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 385    | AlaIleAspProPheAsnSerAspArgMetLeuTyrGly                       | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3126   | CGCGGTGCAACCGG--GATGGTGACGCGCTCAGCTACAGCTGGGAGCAGACGGGTGTTTC  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 402    | rLeuTyrAlaThrAsn  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3183   | GACGGTCCGCATCAATGGCGCCGACAGCTCCGCGCACTCTCTCCGCCACCCCGGTCCCGGG | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 408    | -----As   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3243   | TTCTAGCAGATTCCACGTCGACGCGGACGCGCTCGGCCCTCTCCCTCCAAAGCGGTGCC   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 408    | pLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLys             | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3303   | GGTGACCATCATCGATGTTCTCTGCGCGGCACTCTGCGCGGACGCTGTAACGCGGCAT    | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 427    | uGluGluThrAla--ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSe   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3363   | CGATGCCACGGGAAACGCTGGTGACGCTGGTACGCTGACGCTCAGCGGTCCGCGGAGC    | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 446    | rAlaLeuGlyAsp   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3417   | CGCCGAGGTGACACGCTGACCTACCATCGGAGCAGATCGTGGCAGCGATGTACCTC      | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 454    | eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3477   | GACGGCGCGCGAC--ACCTTGACGCGCTCG--TTCCCGCGCGGACGACGCGGCAG       | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 474    | rGlyThrSerValAspTyrAlaGluLeuAsnProSer                         | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3528   | CGGTACCACTCGGCTTCATC--CTGACGGTCAGCGATGGCAGCTTCGACACCGACGCGA   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 487    | -----IleIleValAlaGlySerPheAspProSerSerGlnProAsnAspAr          | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3585   | TGCGGTGCGCGTCATCGTGGCGGCG--GACCGGGTGCACACCGGAGC--             | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 503    | gHisValAlaPheSerThrAspGlyLysAsnTrpPheGlnGlySerGluProGlyG      | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3631   | -AACACTGCTCCGGAAGCGGACGCGGGCGGAGTCGGCATCTGTG--GCGGAGGGCGGCAC  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 523    | yValThrThrGlyThrValAlaAlaSerAlaAspGlySerArg                   | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3687   | TGTGACGCTCAACGGGACG--GCGACCGACGCTGACGCTGACACGCTCGTCACTGCTCTG  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 541    | pAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAl  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |
| 3744   | GACGAG--ATCGGGGCGGACGCCCGTG--ACGCTGTCTGACGCGCTCTCGCTGAC       | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Db |
| 561    | aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysTh  | 4287   | 175.50 | 35.3%              | 22.8%                 | 47          | Qy |

Db 3795 GCCGAGCTTACCGCCCGGCTCTCTCC-----GATCCGCTGAC 3833  
 Qy 581 rPheTyrAlaLeu---SerAsnGly-----ThrPheTyrArgSerThrAspGlyGlyVa 598  
 Db 3834 CTTCTCTCTGATGTCAGCGAGCGGTACGCGCATGTCTGTCGACGTGACGTCCATCAGCGT 3893  
 Qy 598 lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHi 618  
 Db 3894 GACCAGGAGAGACTCGCG-----CCGGTGGCTCCGCTCGCGGGTGCTC----- 3939  
 Qy 618 sAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSe 638  
 Db 3940 -----TCTGCAACACGACGCTCGGCCACGCTCGACGCGCTCGGCTTCACG 3983  
 Qy 638 rThrAsnGlyGlySer-----SerTrpSerAlaAlaThrGlyValSerSerAl 654  
 Db 3984 CCACCGAAACGGCGATGTCTGACGTACCGCTGGACGCGAGTCTCCGCGCCGACGCGAC 4043  
 Qy 654 aValAsn 656  
 Db 4044 CATCTCC 4050  
 RESULT 15  
 ; Sequence 1119, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; CURRENT FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 1119  
 ; LENGTH: 16584  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 ; US-09-902-540-1119  
 Alignment Scores:  
 Pred. No.: 0.000242 Length: 16584  
 Score: 175.50 Matches: 165  
 Percent Similarity: 35.3% Conservative: 90  
 Best Local Similarity: 22.8% Mismatches: 240  
 Query Match: 4.3% Indels: 229  
 DB: 3 Gaps: 47  
 US-09-917-376-3 (1-740) x US-09-902-540-1119 (1-16584)  
 Qy 24 PheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArg 43  
 Db 6448 TACAACCTCAGTTCCTCCGGGATGCTCCCTCGACCATCGACCTGGCGACGCGCTACGCC 6507  
 Qy 44 TrpAlaAlaAsnGlyArgTrpIleProLeuLeuAsp-----TrpValGly 59  
 Db 6508 GCGAAGACGGTGCAGGTCCGCTTCGGGTGCGTACGACGAGAGCGCGGCTACACGCGC 6567  
 Qy 60 Trp-----AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsn 77  
 Db 6568 TGGCTGTGACGACCTCGAGTTTCAACGGCATC----- 6600  
 Qy 78 ThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly 97  
 Db 6601 ACGAACHACTCGTTCGCGCACCATCGCG-----CCGAGGATGGC 6639  
 Qy 98 AlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLys 117  
 Db 6640 GTTTGT-----GTGAACCCCTTGCCCCATCGCC 6666

Qy 118 LeuGlyGlyAsnMetProGlyArgGlyMet-----GlyGluArgLeu----- 131  
 Db 6667 AACGCGGT-----CCGACCGGTCCATCTCGCGCGGGTAACTGGTGAGCCTTACGGT 6720  
 Qy 132 ---AlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGly 150  
 Db 6721 TCGGCTGCGGATCCGGAAGCCGCGGTGACCTTC----- 6756  
 Qy 151 LeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspVal 170  
 Db 6757 -----ACCTGGGACAG-----ACGCTGCGCCCTGCGGTC 6786  
 Qy 171 -----GlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIle 187  
 Db 6787 ACCCTGGGGGTGCCAACACCCCTGAACCGCTGCTTACGCGCGCGCAGGTTCACGAGTCC 6846  
 Qy 188 GlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLys 207  
 Db 6847 ACCGCACTGGTGTTCACGCTG-----ACCGTCTCCGACGCGGTCAAGACGTCACGCGAC 6900  
 Qy 208 ThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGly 227  
 Db 6901 TCGGTGACGCTGACCGGTGGCGCTCCCAACAATCCG----- 6936  
 Qy 228 AlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPhe 247  
 Db 6937 -----CCACCGTGAACGCGGCGCTCGACGCGCATCGTCGAGGAGCGCGTGTAGTAC 6987  
 Qy 248 AspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly 267  
 Db 6988 -----ACGCTGACGCGGTCTCCGCCAGCGATGCGGATGCG 7020  
 Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIleSerPro 287  
 Db 7021 AATGCC-----CTCAGTACTGTGGACCCAGGTCTCCGGT 7056  
 Qy 288 ValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArg 307  
 Db 7057 ACGCGGTG-----GCGGTGNAGACTAC----- 7080  
 Qy 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIle 327  
 Db 7081 ACCACGCGGCGGACGCTTCATCGCGCGGAGGTACAG-----CTGGATCAGTCTCGTGGTC 7137  
 Qy 328 PheArgSerThr-----AspGlyGlyAlaThrTrp-ThrArgIleTrpAspTrpThrse 345  
 Db 7138 TTCCGTCTGACGCTGACGCGCATCGCATCGCGTGAACG-----ACACG 7182  
 Qy 345 rTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPh 365  
 Db 7183 GTCACCGTGACGCTCACCAACCGCA-----ACCGCGCGCCCA----- 7219  
 Qy 365 eGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMe 385  
 Db 7220 -----TCGTCTCGACACGCTCGGTGCTTCGCGCGGGCACT-GTCACCGTGACGCGCTC 7274  
 Qy 385 tAlaIleAspProPheAsnSerAspArgMetLeuTyrGly-----ThrGlyAlaTh 402  
 Db 7275 CGCGTTCGACCCG-----GATGGTGACGCTCAGCTACAGCTGGGAGCAGACGCGGTGTC 7331  
 Qy 402 rLeuTyrAlaThrAsn----- 407  
 Db 7332 GACGTCGCCATCAATGGCGCGACACGTCGCCCATCTCTTTCGCCACCCCGGTCCCGGG 7391  
 Qy 408 -----As 408  
 Db 7392 TTCGTACAGTTTACCGTGACGCGGACGCGGCTCGGCTCTGCTCTCAAGCGGTGTC 7451  
 Qy 408 pLeuThrLysTrpAspSerGlyGlnIleHisIleAlaProMetValLys---GlyLe 427  
 Db 7452 GGTGACCATCATCGATGTTCTCTCGCGCGCAACTCTGCGCGCGACGCTGAACGCGGCGCAT 7511

```
QY 427 uGluGluThrAla---ValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSe 446
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7512 CGATGCCACGCGGAACGCTGGTGACGTCGTGACGCTCAGCGCTCCGCGAGC-----GA 7565

QY 446 rAlaLeuGlyAsp-----LeuGlyGly
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7566 CGCCGAGGGTGACAGCTGACGTACCACTGGGACGACATCGTGGCAGCGATGTGACCCT 7625

QY 454 eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThrTh 474
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7626 GACGGGCGCGAC---ACCCTGACGCGCTCG-----TTACCGCGCGGACGACGCGCAG 7676

QY 474 rGlyThrSerValAspTyrAlaGluLeuAsnProSer----- 486
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7677 CGGTACCACGCTCGGCTTCATC---CTGACGGTCAGCGATGGCACCTCGACCCAGCGA 7733

QY 487 -----IleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspAr 503
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7734 TGCGTCGGGTCACTGTGGCGG-----GACCGGGTCCCAACCCGAGC----- 7779

QY 503 gHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGl 523
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7780 -AACACTGCTCCGGAAGCGAGCGCGGAGTCGGCCATCGTG---GCGAGGGCGGCC 7835

QY 523 yValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg-----PheValTr 541
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7836 TGTGACGCTCAACGGCAG---GCGACCGACGCTGACGGTGACACGCTCGTCACTCGTCTG 7892

QY 541 pAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAl 561
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7893 GAGCGAG---ATCGGCGGCGACGCCCGTG-----ACGCTGTCTGACGCGCTCTCGCTGAC 7943

QY 561 aAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysTh 581
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7944 GCCGACGTTACCGCCCGGCTCTCTC-----GATCCGCTGAC 7982

QY 581 rPheTyrAlaLeu---SerAsnGly-----ThrPheTyrArgSerThrAspGlyGlyVa 598
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7983 CTTCTCTCTGATGTCAGCGACGTACGGGACTGTCTGTCACGTGACGTCCATCAGCGT 8042

QY 598 lThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHi 618
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8043 GACCGAGGAGAACGTCGCG-----CCGGTGGCGTCCGCTCGCGGTGCTC----- 8088

QY 618 sAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSe 638
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8089 -----TCTGGCAACGACGTCGCGCCAGCTCGACGCTCGGCTCGGCTTCCAG 8132

QY 638 rThrAsnGlyGlySer-----SerTrpSerAlaIleThrGlyValSerSerAl 654
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8133 CGACGGAACGGCGATGTGTGACGTACCGTGGACCGAGGTCTCCGGCCCGAACGGGAC 8192

QY 654 aValAsn 656
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8193 CATCTCC 8199
```

Search completed: March 2, 2006, 13:50:24  
Job time : 5618.94 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:19:26 ; Search time 1388.78 Seconds  
(without alignments)  
4406.259 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPTWSNAIVGGGQFVD.....YGTNGRGIVGIGGAPSG 740

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abss/ABSSWEB spool/US09917376/runat\_02032006\_091500\_8223/app\_query.fasta\_1  
-DB=Published Applications\_NA\_Main -QMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss06p  
-USER=US09917376 @CGN\_1\_1\_3095 @runat\_02032006\_091500\_8223 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA\_Main:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score  | Query Match | Length  | ID | Description        |
|------------|--------|-------------|---------|----|--------------------|
| 1          | 4036   | 100.0       | 2869    | 3  | US-09-917-376-2    |
| 2          | 4036   | 100.0       | 2869    | 5  | US-10-155-400-2    |
| 3          | 2429.5 | 60.2        | 2646    | 6  | US-10-156-761-1845 |
| 4          | 2429.5 | 60.2        | 9025608 | 6  | US-10-156-761-1    |
| 5          | 1579   | 39.1        | 2517    | 5  | US-10-026-994-4    |
| 6          | 1579   | 39.1        | 2710    | 5  | US-10-026-994-1    |
| 7          | 1405   | 34.8        | 2217    | 6  | US-10-156-761-2561 |
|            |        |             |         |    | Sequence 2, Appli  |
|            |        |             |         |    | Sequence 2, Appli  |
|            |        |             |         |    | Sequence 1845, Ap  |
|            |        |             |         |    | Sequence 1, Appli  |
|            |        |             |         |    | Sequence 4, Appli  |
|            |        |             |         |    | Sequence 1, Appli  |
|            |        |             |         |    | Sequence 2561, Ap  |

Alignment Scores:  
Pred. No.: 0  
Score: 4036.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Length: 2869  
Matches: 740  
Conservative: 0  
Mismatches: 0

## ALIGNMENTS

RESULT 1

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; TITLE OF INVENTION: CELLULOGLYTICUS

; FILE REFERENCE: 40197.4US01

; CURRENT APPLICATION NUMBER: US/09/917,376

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 2

; LENGTH: 2869

; TYPE: DNA

; ORGANISM: Acidothermus cellulolyticus

; FEATURE:

; NAME/KEY: modified\_base

; LOCATION: (2869)

; OTHER INFORMATION: a, c, t, g, other or unknown

US-09-917-376-2

|   |    |       |      |         |    |                      |                    |
|---|----|-------|------|---------|----|----------------------|--------------------|
| c | 8  | 1405  | 34.8 | 9025608 | 6  | US-10-156-761-1      | Sequence 1, Appli  |
|   | 9  | 1139  | 28.2 | 3668    | 3  | US-09-927-827-21     | Sequence 21, Appli |
|   | 10 | 1117  | 27.7 | 2367    | 7  | US-10-395-241-13     | Sequence 13, Appli |
|   | 11 | 1117  | 27.7 | 2646    | 7  | US-10-395-241-11     | Sequence 11, Appli |
|   | 12 | 1114  | 27.6 | 2481    | 7  | US-10-395-241-17     | Sequence 17, Appli |
|   | 13 | 878.5 | 21.8 | 5698    | 7  | US-10-420-191-1      | Sequence 1, Appli  |
|   | 14 | 392.5 | 9.7  | 1103    | 8  | US-10-653-047-7511   | Sequence 7511, Ap  |
|   | 15 | 238   | 5.9  | 7407    | 6  | US-10-246-330-3      | Sequence 3, Appli  |
|   | 16 | 238   | 5.9  | 7407    | 7  | US-10-282-122A-30151 | Sequence 30151, A  |
|   | 17 | 199.5 | 4.9  | 7305    | 7  | US-10-282-122A-11269 | Sequence 11269, A  |
|   | 18 | 195   | 4.8  | 5295    | 7  | US-10-282-122A-11871 | Sequence 11871, A  |
|   | 19 | 192   | 4.8  | 9903    | 7  | US-10-282-122A-28185 | Sequence 28185, A  |
|   | 20 | 188.5 | 4.7  | 4399    | 6  | US-10-369-493-37642  | Sequence 37642, A  |
|   | 21 | 184   | 4.6  | 3324    | 7  | US-10-282-122A-26288 | Sequence 26288, A  |
|   | 22 | 182   | 4.5  | 69350   | 7  | US-10-041-018-379    | Sequence 379, App  |
|   | 23 | 180   | 4.5  | 2355    | 6  | US-10-369-493-37792  | Sequence 37792, A  |
|   | 24 | 179.5 | 4.4  | 3222    | 7  | US-10-282-122A-14432 | Sequence 14432, A  |
|   | 25 | 176   | 4.4  | 3267    | 5  | US-10-158-160A-14    | Sequence 14, Appli |
|   | 26 | 174   | 4.3  | 2855    | 10 | US-11-097-143-40748  | Sequence 40748, A  |
|   | 27 | 174   | 4.3  | 5221    | 10 | US-11-097-143-26683  | Sequence 26683, A  |
|   | 28 | 174   | 4.3  | 5302    | 10 | US-11-097-143-40747  | Sequence 40747, A  |
|   | 29 | 173   | 4.3  | 2708    | 10 | US-11-097-143-26684  | Sequence 26684, A  |
|   | 30 | 170.5 | 4.2  | 2176    | 10 | US-11-097-143-14408  | Sequence 14408, A  |
|   | 31 | 170.5 | 4.2  | 4252    | 10 | US-11-097-143-14407  | Sequence 14407, A  |
|   | 32 | 170.5 | 4.2  | 7191    | 7  | US-10-282-122A-35048 | Sequence 35048, A  |
|   | 33 | 170.5 | 4.2  | 25165   | 5  | US-10-114-170-39     | Sequence 39, Appli |
|   | 34 | 170   | 4.2  | 4323    | 7  | US-10-282-122A-8383  | Sequence 8383, Ap  |
|   | 35 | 169.5 | 4.2  | 3360    | 6  | US-10-245-802-11     | Sequence 11, Appli |
|   | 36 | 169   | 4.2  | 69350   | 7  | US-10-041-018-379    | Sequence 379, App  |
|   | 37 | 168   | 4.2  | 5451    | 6  | US-10-101-510-597    | Sequence 597, App  |
|   | 38 | 167.5 | 4.2  | 2110    | 5  | US-10-027-632-97984  | Sequence 97984, A  |
|   | 39 | 167.5 | 4.2  | 2110    | 6  | US-10-027-632-97984  | Sequence 97984, A  |
|   | 40 | 167   | 4.1  | 1839    | 6  | US-10-369-493-34868  | Sequence 34868, A  |
|   | 41 | 166   | 4.1  | 3879    | 7  | US-10-282-122A-8717  | Sequence 8717, Ap  |
|   | 42 | 165.5 | 4.1  | 4383    | 6  | US-10-245-802-21     | Sequence 21, Appli |
|   | 43 | 165   | 4.1  | 6615    | 7  | US-10-282-122A-28180 | Sequence 28180, A  |
|   | 44 | 165   | 4.1  | 67167   | 9  | US-10-937-730A-3     | Sequence 3, Appli  |
|   | 45 | 164.5 | 4.1  | 2540    | 9  | US-10-680-860A-161   | Sequence 161, App  |



|  |        |  |      |
|--|--------|--|------|
| Query Match:                                       | 100.0% | Indels:  | 0    |
| DB:  | 3      | Gaps:  | 0    |
| US-09-917-376-3 (1-740) x US-09-917-376-2 (1-2869) |        |  |      |
| QY   | 1      | AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp      | 20   |
| DB   | 139    | CGGACGACTCAGCGGTACACCTGGAGCAACGTGGGATCGGGGGCGGCTTTGTGCAC       | 198  |
| QY   | 21     | GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly   | 40   |
| DB   | 199    | GGGATCGCTTCAATGAAGTGCACCGGAATCTGTACGTGGAGCGACATCGGGGG          | 258  |
| QY   | 41     | MetTyrArgTrpAspAlaIleAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp   | 60   |
| DB   | 259    | ATGTATCGATGGGATCGCGCAACCGGGGTGATCCCTCTTCTGGATTGGTGGATGG        | 318  |
| QY   | 61     | AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaIleAspProIleAsnThrAsnLys   | 80   |
| DB   | 319    | AACAAATGGGGGTACACCGCGTGTGAGCAATTCGGCGAGACCCGATCAATCAACAAG      | 378  |
| QY   | 81     | ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu   | 100  |
| DB   | 379    | GTATGGCGCGCGTGGGAATGTACACCAACAGCTGGGACCCCAACCGAGCGATTCTC       | 438  |
| QY   | 101    | ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly   | 120  |
| DB   | 439    | CGCTCTGTCTGATCAGGGCGCAACGTGGCAATTAACGCCCTTGCCTTCAAGCTTGGCGG    | 498  |
| QY   | 121    | AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle   | 140  |
| DB   | 499    | AACATCCCGGGCGTGGAAATGGCGAGCGGCTTGGGTGGATCCCAACCAATGACACATT     | 558  |
| QY   | 141    | LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr   | 160  |
| DB   | 559    | CTGTATTTTGGCGCCCGAGCGCAAGGGCTCTGGAGAGCACACAGATTCGGCGCGAAC      | 618  |
| QY   | 161    | TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr   | 180  |
| DB   | 619    | TGCTCCAGATGACGAATTTTCGGACGTAGGCACGTACATTGCAATCCCACTGACACG      | 678  |
| QY   | 181    | ThrGlyTyrGlnSerAspIleGlnIleValTrpValAlaPheAspLysSerSerSer      | 200  |
| DB   | 679    | ACCGGTATTCAGAGCGATATTCAAGCGCTGCTGGGTGCGCTTTCGACAGTCTTCGTCA     | 738  |
| QY   | 201    | SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal   | 220  |
| DB   | 739    | TGCTCGGCAAGCGAGTAAAGACCAATTTTGTGGCGTGGCGATCCCAATAATCCGGTC      | 798  |
| QY   | 221    | PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe   | 240  |
| DB   | 799    | TTCTGGAGCAGAGACGCGCGGCGACGTGGCAGCGGTGCGGGTGGCGGCGCTTC          | 858  |
| QY   | 241    | IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn   | 260  |
| DB   | 859    | ATCCCGCACAAAGGGCTCTTTGACCCGGTCAACACCGTCTATATTGCCACACCAAT       | 918  |
| QY   | 261    | ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly   | 280  |
| DB   | 919    | ACGGGTGGTCCGATACCGGAGCTCCGGCAGCTGTGAAATTCCTCGGTGACCTCCGGG      | 978  |
| QY   | 281    | ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr   | 300  |
| DB   | 979    | ACATGGACCGGAATACGCCCGGTACCTTCGACGGACACCGGCCAACGACTACTTTGGTTAC  | 1038 |
| QY   | 301    | SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer   | 320  |
| DB   | 1039   | AGCGGCTCACTATGACCGCGCAGCACCCGAAACAGATATGGTGGCAACCCAGATATCG     | 1098 |
| QY   | 321    | TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle   | 340  |
| DB   | 1099   | TGGTGGCGGACCACTAATCTTTCGGAGCACCGACGGCGGTGGAGCTGGACCGGATC       | 1158 |
| QY   | 341    | TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu   | 360  |
| DB   | 1159   | TGGATTGGACGAGTATCCCAATTCGAAGCTTGGCATATGCTGACATTTCCGCGGAG       | 1218 |
| QY   | 361    | ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp      | 380  |
| DB   | 1219   | CCTTGGCTGACCTTCGGGTACAGCCGAACTCTCCGTACCAAGTCCGAGTCCGAGCTCGCTGG | 1278 |
| QY   | 381    | MetAspGluAlaMetAlaIleAspProPheAsnSerAspMetLeuTyrGlyThrGly      | 400  |
| DB   | 1279   | ATGGATCAAGCGATGGCAATCGATCCGTCAACTCTGATCGATGCTCTACGGAACAGC      | 1338 |
| QY   | 401    | AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle   | 420  |
| DB   | 1339   | GCAGCTGTGTACGCAACAAATGATCTCACGAAGTGGGACTCCGGCGGCGACATTCATATC   | 1398 |
| QY   | 421    | AlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuLysSerProSer      | 440  |
| DB   | 1399   | GCCTCGATGTCAAGGATGGAGAGACGGCGTAAACGATCTCATCAGCCCGCGCTCT        | 1458 |
| QY   | 441    | GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThr   | 460  |
| DB   | 1459   | GGCGCCCGCTCATCAGGCTCTCGGAGACCTCGGGGCTTCACCAGCCCGACGTTACT       | 1518 |
| QY   | 461    | AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr   | 480  |
| DB   | 1519   | GGCGTGCATCGACGATCTTCACTCACCGTGTTCACAGCCGCGCACCGCTGCTGCTAT      | 1578 |
| QY   | 481    | AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro   | 500  |
| DB   | 1579   | GCAGATTTGAATCCGTGCATCATCTTCGGCTCGAAGTTTCGATCCATCGAGCCAAACG     | 1638 |
| QY   | 501    | AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu   | 520  |
| DB   | 1639   | AACGACAGGCGACGTGCGGTCTCGACAGAGCGGCGCAAGACTGGTTCAGAGGACGCGA     | 1698 |
| QY   | 521    | ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal   | 540  |
| DB   | 1699   | CCTGGCGGGGTGACGACGGCGGCGCACGTCGCGCATCGGCGGACGGCTCTCGTTCTGTC    | 1758 |
| QY   | 541    | TrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrp   | 560  |
| DB   | 1759   | TGGGCTCCCGCGCATCCCGTTCAGCTGTGTGTGTACGACGTGGATTTGGCACTCTCTG     | 1818 |
| QY   | 561    | AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys   | 580  |
| DB   | 1819   | GCTGCTTCGAAGGTGTTCGCCCAATGCCAGATCCGCTCAGACCGGGGTGAATCCAAAG     | 1878 |
| QY   | 581    | ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe   | 600  |
| DB   | 1879   | ACTTTCTATGCCCTATCCATGGACCTTCTATCGAAGCACGAGCGGGGTGACATTC        | 1938 |
| QY   | 601    | GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal   | 620  |
| DB   | 1939   | CAACCGGTCCGCGCGGTCTTCCGAGCAGCGGTGCGCTCGGTGATGTTCCACCGGTG       | 1998 |
| QY   | 621    | ProGlyLysGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn      | 640  |
| DB   | 1999   | CCTGGAAGAAAGCGGATCTGTGGCTCGCTGCAATCGAGCGGGCTTTTACCACTCAACCAAT  | 2058 |
| QY   | 641    | GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly   | 660  |
| DB   | 2059   | GGCGGACGAGTGTGCTGCAATCACCGGGTATCTTCGCGGTGACGTTGGATTGGT         | 2118 |
| QY   | 661    | LysSerAlaProGlySerSerTrpProAlaValPheValValGlyThrIleGlyGlyVal   | 680  |
| DB   | 2119   | AGTCTGGCGCGGTGCTCATACCGAGCGTCTTTGTCGTCGCGACGATCGGAGGCGTT       | 2178 |
| QY   | 681    | ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspGln      | 700  |
| DB   | 2179   | ACGGGGCGGTACCGCTCCGACGACTGTGGGACGACCTGGGTACTGATCAATGATGACGAG   | 2238 |
| QY   | 701    | HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal      | 720  |

Db 2239 CACCANTACGGAATTTGGGACACAGCAATCACCGGTGACCAAGCAATTTACGGCGGGTG 2298  
Qy 721 TyrileGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly 740  
Db 2299 TACATAGGACGACGCGCGTGGAAATTCATACGGGAGCATTTGGTGGTGGCGCGTCCGGA 2358

## RESULT 2

US-10-155-400-2  
; Sequence 2, Application US/10155400  
; Publication NO. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2869  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (2869)  
; OTHER INFORMATION: a, c, t, g, other or unknown  
US-10-155-400-2

## Alignment Scores:

Pred. No.: 0 Length: 2869  
Score: 4036.00 Matches: 740  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 5 Gaps: 0

US-09-917-376-3 (1-740) x US-10-155-400-2 (1-2869)

Qy 1 AlaThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyPheValAsp 20  
Db 139 GCGAGCTACCGCTACACCTTGAGCAACGTGGCGATCGGGCGCGCGCTTTGTCCAC 198  
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGly 40  
Db 199 GGGATCGTCTTCAATGAAGGTGCACCGGAATTCGTACGTGCGGACGACATCGGGGG 258  
Qy 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrp 60  
Db 259 ATGTATCGATGGGATGCGCGCAACGCGCGGTGGATCCCTCTTCGTGATGGGTGGGATG 318  
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIys 80  
Db 319 AACAAATGGGGGTACAACGGCGTGTACAGATGCGGACACCCCGATCAATACATAACAG 378  
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100  
Db 379 GTATGGCGCGCGTCCGAAATGTACACCAACAGCTGGGACCCCAACAGACGGAGCATTC 438  
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly 120  
Db 439 CGCTCGTCTGATACGGGCGCAACGTGGCAATAACGCCCTTCCGCTTCAAGCTTGGCGGC 498  
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIle 140  
Db 499 AACATGCGCGCGCGTGGAAATGGGCGAGCGGTTCGGTGGATCCAAAACAAATGACAACAT 558  
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160

Db 559 CTGTATTTCCGGCGCCCGAGCGCAAGGGCTCTGGAGAAGCACAGATTCCGGCGCGACC 618  
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180  
Db 619 TGGTCCAGATGACGAATTTTCGGACGTAGGACAGTACATTGCAATCCCACTGACACG 678  
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 200  
Db 679 ACCGGCTATCAGAGCGATATTCAAGGGCTGCTGGGTGCGCTTTCCGCAAGTCTTCGTCA 738  
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220  
Db 739 TCCTCGGCAAGCGAGTAGACCAATTTTGTGGCGGTGGCGATCCCAATAATCCCGTC 798  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 799 TTCTGGAGCAGAGACGCGCGCGCGAGCGTGGCAGCGGTGCGGGTGGCGACCGGCTTC 858  
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260  
Db 859 ATCCCGCACAAAGGGCGTCTTTGACCGGTCAACACCGTCTCTATATTGCCACCAAGCAAT 918  
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280  
Db 919 ACGGGTGGTCCGTATCAGCGGAGCTCCGGCAGCTCTGGAAATTCCTCGTGACCTCCGG 978  
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300  
Db 979 ACATGGACCGGAATCAGCCCGGTACCTTTCGAGCAGCACCGGCAACGACTACTTTGGTTAC 1038  
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320  
Db 1039 AGCGGCTCTCATTCGACCGCAGCACCGCATATGTTGGCAACCCAGATATCG 1098  
Qy 321 TrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340  
Db 1099 TGGTGGCGCGACACCAATAATCTTTTCGAGCAGCCGCGGTGCGAGTGGACGCGGATC 1158  
Qy 341 TrpAspTrpTrpSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360  
Db 1159 TGGGATTGGACGAGTTATCCCAATCGAAGCTTCGGATATGTTGTGACATTTCCGGCGAG 1218  
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrp 380  
Db 1219 CTTTGGCTGACCTTCGGCGTACAGCCGATCTCTCCGTACCCAGTCCGAGCTCGGCTGG 1278  
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400  
Db 1279 ATGGATGAAGCGATGCGCAATCGATCGTTCAACTCTGTATCGGATGCTCTACGGAACAGGC 1338  
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420  
Db 1339 GCGAGCTTGTACGCAACAAATGATCTCAGAGTGGGACTCCGGCGCGCAGATTATCATC 1398  
Qy 421 AlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProProSer 440  
Db 1399 GCGCGATGTTCAAGGATGGAGGAGCGCGGTAACGATCTCATCAGCCCCCGCT 1458  
Qy 441 GlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPheThrHisAlaAspValThr 460  
Db 1459 GCGCGCGCTCATCAGCGCTCTCGAGACCTCGCGGCTTCACCCAGCGCGAGTTACT 1518  
Qy 461 AlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyr 480  
Db 1519 GCGGTGCCATCAGCATCTTCAGTCACCGGTGTTCCAGCACCGGACCGCGTCCGACTAT 1578  
Qy 481 AlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSerSerGlnPro 500  
Db 1579 GCGGAATTTGAATCCGTTCGATCATCGTTCGCGCTGGAAATTTTCGATTCATCGAGCAACCG 1638  
Qy 501 AsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGlu 520  
Db 1639 AACGACAGCAGCTCCGTTCTCGACAGACGCGCGGCAAGACTGGTTTCCAAAGGACGGA 1698

QY 521 ProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheVal 540  
Db 1699 CTTGGCGGGTACGACGACGGCGGCACCGTCCGCGCATCGCCGACGCTCTCGTTTCGTC 1758  
QY 541 TrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrp 560  
Db 1759 TGGGCTCCCGCGCATCCCGTACGCTGCTGGTGTACGACGCTGGATTGGCAACTCTCTGG 1818  
QY 561 AlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLys 580  
Db 1819 GCTGCTTCGAAGTGTTCGCGCCATGCCAGATCCGCTCAGACCGGGTGATCCAAAG 1878  
QY 581 ThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPhe 600  
Db 1879 ACTTTCTATGCCCTATCCATGAACCTTCTATCGAAGCACGACGCGCGGTGACATTC 1938  
QY 601 GlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaVal 620  
Db 1939 CAACCGCTCCGCGCGCTCTTCCGACGACGCGGTGCCGTGCTCATGTTCCACGCGGTG 1998  
QY 621 ProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsn 640  
Db 1999 CTGGAAAGAGGCCATCTGTGGCTCGCTGCATCGAGCGGGCTTTACCACTCAACCAAT 2058  
QY 641 GlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGly 660  
Db 2059 GCGCGCAGCAGTTGGTCTGCAATCACCGCGGTATCTCCGCGTGAACGGGTATGGT 2118  
QY 661 LysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyVal 680  
Db 2119 AAGTCTCGCCCGCGGTGTCATACCCAGCGCTTTGTGCTCGGCACGATCGGAGCGTT 2178  
QY 681 ThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValleuIleAsnAspAsnGln 700  
Db 2179 ACGGGCGGTACCGCTCCGACGACTGTGGACACCTGGGTACTGATCAATGATGACCAG 2238  
QY 701 HisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgVal 720  
Db 2239 CACCAATACGGAATTTGGGACAAAGCAATCACCGGTGACCACCGAATTTACGGCGGTG 2298  
QY 721 TyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaProSerGly 740  
Db 2299 TACATAGGCACGAACGCGCGTGAATTTGATACGGGACATTTGTTGTCGCGCTCCGGA 2358

## RESULT 3

US-10-156-761-1845  
; Sequence 1845, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHERA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1845  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2646)

## US-10-156-761-1845

Alignment Scores:  
Pred. No.: 5,87e-224 Length: 2646  
Score: 2429, 50 Matches: 438  
Percent Similarity: 74.0% Conservative: 111  
Best Local Similarity: 59.0% Mismatches: 174  
Query Match: 60.2% Indels: 19  
DB: 6 Gaps: 8

US-09-917-376-3 (1-740) x US-10-156-761-1845 (1-2646)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 115 TACAGCTGGAAGAACGCGCGTTCGACGCGCGCGCTTCGTCGCCGCACTCGTCTTCAAC 174  
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 175 CGCTCCGAGAAACCTTCGCTACCGCGCACCGACATCGCGCGCGCTTACCGCTGGGCC 234  
QY 46 AlaAlaAsnGlyArgTrpIleProleuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 235 GAGTCTCGAAGACTGAGCGCGCTGCTCGACTCGTTCGCTCGGCTGAGGACACTGGGGGCAC 294  
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 295 ACGGTTGTGTCGAGCCCTCCGACTCCGCTCGACCCGACCAAGGTGTACGCGCGCGTC 354  
QY 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 355 GGACGCTACAGAACAGCTGGGACCGCGGCAACCGTGCCTGCTCAGGTTCGCGGACCGG 414  
QY 106 GlyAlaThrTrpGlnIleThrProleuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 415 GCGCGAGCTGGCAGAACCGACCTGCCCTTCAAGCTGGCGGGAACATGCGCGCGCGG 474  
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 475 GGCATGGCGGAGCGGCTCGCGGTTCGACCCGAAACAGGACAGCGTCTGTATCTCGCGCGG 534  
QY 146 ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
Db 535 CCCAGCGCAAGGGGCTGTGCGGTTCGACGACCTCGGGGCTCTCTGTCGCGAGGTCAAC 594  
QY 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
Db 595 GACTTCCCGAACGTCGGCACCTACGTGCAGACCGCACCGACACGCGGGGTACGCGCTCC 654  
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla 205  
Db 655 GACAAACAGGCGCATCTGTGGTTCACCTTCGACGAGTTCGACGGGTTCGCGGGAGCTCC 714  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225  
Db 715 ACGCGGACGGTGTACGTGGGTTCGCGCAAGGACAACTCCGTCTATCGCTCCACGGAC 774  
QY 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245  
Db 775 GCGGCGCGACCTGTGTCGCGCGCGCACCGCCCATCTCGCCCAAGGGC 834  
QY 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265  
Db 835 GTGCTGAGCGGCGGAACCGGTGTCTGTACCTTCGGTACAGGCGCGGACCGGTAC 894  
QY 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle 285  
Db 895 GACGCGCGCAAGGACAGCTGTGGGTACACGACAGACCGGGACCTGGAGCAACATC 954  
QY 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIle 305  
Db 955 AGCCCGGTTCGGAGGCGGACAC-----TACTACGGGCTTCAGCGGGGTGACCGGTG 1005  
QY 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr 325

Db 1006 GACCGGACATCCGGGACGGTGTGCGACTCGGTACAGCTCTCTGGTGGCCGACACG 1065  
Qy 326 TlellePheArgSerThrAspGlyGlyAlaThrThrArgIleTTPAspTrpThrSer 345  
Db 1066 CAGCTCTTCGCTCCACGACAGCGCGGCGACCTTGGACGAAGCGCTGGACTACACCTCG 1125  
Qy 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe 365  
Db 1126 TATCCGAGCCGCTCGAACCGCTTCACCATGGATGCTCTGCTCCGCTCGCTCACCTCG 1185  
Qy 366 GlyValGlnProAsnProValProSerProLysLeuGlyTTPMetAspGluAlaMet 385  
Db 1186 GGAGGAAACCCCGCACCGCCCGACGACAGACCCCGAACTCGGCTGGATGACCGAGTCCCTG 1245  
Qy 386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 1246 GAGATCGACCGCTTCGACTCGCGCGCATGATGACGAAGCGGCGGCGCATCGCTACCGC 1305  
Qy 406 ThrAsnAspLeuThrLysTTPAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
Db 1306 ACGGACAACTTGCAGAACTGGGACAGCGGAGCCAGTTTACCATCAAGCCGATGGCGCG 1365  
Qy 426 GlyLeuGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444  
Db 1366 GGCCTGGAGGAGACGGCGCTCAACGAOCTCCCTCGCTCCGCGGCGGCGCCAGCTG 1425  
Qy 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
Db 1426 TTCAGCGGCTCGGTGATCATCGCGGCTTCGGGACACGGACCTCACACCGGTCCGCTG 1485  
Qy 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
Db 1486 CTGATGTACAGCTCGCGGAATTCACACGAGCACCGCTCGACTAGCCGCGAGACCGAC 1545  
Qy 485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504  
Db 1546 CCGGCGACGGTGTGGCGGTTCGGAATCTCGAC-----TCGGGTCCG-----CAT 1590  
Qy 505 ValAlaPheSerThrAspGlyGlyLeuAsnTrpPheGlnGlySerGluProGlyGlyVal 524  
Db 1591 GTGGCGTTCGACGGAACACGGCGCAACTGGTTTCGGGGCGGACCTTCGGGGGTC 1650  
Qy 525 ThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 544  
Db 1651 AGCGGGGTGGAGCGTTCGGCGCGCTCCGACGCGAGTCTGCTGTGTGGAGCGCGCG 1710  
Qy 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
Db 1711 GGCACCGG-----GTGCGATCACGACCGGGTTTCGGCACCTCGTGTGGGTTCGCG 1764  
Qy 565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584  
Db 1765 GGCCTCCGCGCGGGGGGATGTCGAGTCCGACCGGGTTCGACCCGGAAGACCTTTCAGCG 1824  
Qy 585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604  
Db 1825 TTCAGTCCGCGAGTTCAGTTCAGTTCGACGCGGGCGGACCTTCACGGGTTCGCG 1884  
Qy 605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623  
Db 1885 GCCACGGGCTCCGAGCGGCGACAGC-----GTGCGCTTCAAGGCGCTCCCGGCGACG 1938  
Qy 624 GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer 638  
Db 1939 AAGGCGACATCTGGCTGGCGGGCGGCGGAGCGGCGGACGCGCGGTACGGGCTGTGGCACTCG 1998  
Qy 639 ThrAsnGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 1999 ACGACGCGCGCGCGGCTTCACCAAGCTCCGACCGCTCCGACCGAGCGGCGACCATCGGC 2058  
Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly 678

Db 2059 TTCGCAAGCGGCGACCGCGCTCGTACACAGCGCTCTACACAGCGCGAAGATCGGC 2118  
Qy 679 GlyValThrGlyAlaTyrArgSerAspSerCysGlyThrThrTrpValLeuIleAsnAsp 698  
Db 2119 GGTGTGCGCGCATCTTCGGTCCGACCGACCAAGGCGCGAGCTGGACCGCGTCAACGAC 2178  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2179 GATGCCCACTAGTGGGTGGGCGCGCGCGCATCACCGGTGACCCCGAGGTCTACGGG 2238  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaPro 738  
Db 2239 CGCGTGTATGTCTCGACGAACCGGCGCGGATCGTCTACGCGGACACCGCGGCTCTTCG 2298  
Qy 739 SerGly 740  
Db 2299 GACGGC 2304  
RESULT 4  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1  
Alignment Scores:  
Pred. No.: 1,86e-219 Length: 9025608  
Score: 2429.50 Matches: 438  
Percent Similarity: 74.0% Conservative: 111  
Best Local Similarity: 59.0% Mismatches: 174  
Query Match: 60.2% Indels: 19  
DB: 6 Gaps: 8  
US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608)  
Qy 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 2277508 TACAGTGGAGAGACCGCGCGGTTCGTCACCGCGCGGCTTCGTCGCGCATCGTCTTCAAC 2277567  
Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 2277568 CGCTCCGAGAGAACCTCGCTTACCGCGCACCGCATCGCGCGCGCTTACCGTGGCGC 2277627  
Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65  
Db 2277628 GAGTCTCTGAAGACCTCGACCGCGCTGTCTCGACTCGGTGGAGCGACTGGGGGCGAC 2277687  
Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85

Db 2277688 ACGGGTGTGAGCTCGCTCGACTCCGTGACCCGGAACAGGTGTACCGCGCGCTC 2277747  
Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaLeuArgSerSerAspGln 105  
Db 2277748 GGCACGTACACAGACAGCTGGGCAACCGGCAACCGGTGCTCAGGTCCGGGACCGG 2277807  
Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 2277808 GGGCGAGCTGGCAGAGACGACCTGCCCTTCAAGCTGGCGGGAACATGCGGGCGG 2277867  
Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
Db 2277868 GGCATGGCGGAGCGGCTCGGGTCGACCCGAAACAGGACGCTGCTGTATCTCGGGCGG 2277927  
Qy 146 ProSerGlyLysLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165  
Db 2277928 CCCAGCGCAAGGGGCTGTGGCTGACGAGCTCGGGGGCTCTCTGGTCAGGTCAAC 2277987  
Qy 166 AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer 185  
Db 2277988 GACTTCCCGCAACGTGCGCACCTAGTGCAGGACGCGACCGACAGCGGTACGCTCC 2278047  
Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerSerSerSerSer 205  
Db 2278048 GACACACAGGGCATTCTGTGGTCACTTCAGAGTGCAGGGGTCCCGGGGAGCTCC 2278107  
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArgAsp 225  
Db 2278108 ACGCGGACGGTGTACGTGCGGGTCGCGACAAAGGACAACTCCGTCTATCGTCCACGGAC 2278167  
Qy 226 GlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHisLysGly 245  
Db 2278168 GCGGGCGCGACCTGGTCCGCTGGCGGCGAGCCACCGGCCCATCTCCGCCACAAAGGC 2278227  
Qy 246 ValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr 265  
Db 2278228 GTGCTGACCGCGCAACCGGTGTCTACTCCGCTACGAGCAAGGCGGACCGGTAC 2278287  
Qy 266 AspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgile 285  
Db 2278288 GACGGCGCAAGGACAGCTGTGGGTACACGACGAGACCGGGACCTGGACGAACATC 2278347  
Qy 286 SerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrile 305  
Db 2278348 AGCCCGGTCCGAGGCGGACAC-----TACTACGGCTTCAGCGGCTGACCGGTG 2278398  
Qy 306 AspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr 325  
Db 2278399 GACCGGAGCATCCGGGACCGGTGATGGCGACTGCGTACAGCTCCTGGTGGCGGACAG 2278458  
Qy 326 IleIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgileTrpAspTrpThrSer 345  
Db 2278459 CAGCTCTTCGCTCCAGGACAGCGCGGACCTGGACGAGGCTGGGACTACACCTCG 2278518  
Qy 346 TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPhe 365  
Db 2278519 TATCCGAGCGCTCGAACCGCTTACCATGATGTCTCGTCTCCCTCGCTCAGCTCG 2278578  
Qy 366 GlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGluAlaMet 385  
Db 2278579 GGAGCGAACCCCGCGCGGACGACCCCGAAACTCGGCTGGATGACCGAGTCCCTG 2278638  
Qy 386 AlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAla 405  
Db 2278639 GAGATGACCCGTTCGACTCCGCGCGCATGATGTACGGAACGCGCGGCTACCGG 2278698  
Qy 406 ThrAsnAspThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
Db 2278699 ACGGACAACTGACGAACCTGGACGCGGAGCGGAGCCAGTTCACATCAGCGATGGCGCGG 2278758  
Qy 426 GlyLeuGluThrAlaValAsnAspLeuIleSerProProSer---GlyAlaProLeu 444  
Db 2278759 GGCCTGGAGGAGACGGCGGTCAACGACCTCGCTCGCTCCCTCCGCGCGGCCCGACCTG 2278818

Qy 445 IleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSer 464  
Db 2278819 TTCAGCGCGCTCGGTGACATCGCGGCTTCGGGCACACGGACCTCACACGGTCCGCTCG 2278878  
Qy 465 ThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsn 484  
Db 2278879 CTGATGTACAGTCCGCCAACTTCCACACGACACCAACGCTCGACTACGCCGAGACCGAC 2278938  
Qy 485 ProSerIleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHis 504  
Db 2278939 CCGGCACGGTGGTGGGTCCGCAATCTCGAC-----TCGGGTCCG-----CAT 2278983  
Qy 505 ValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyVal 524  
Db 2278984 GTGGGCTTCTCGACGGAACACCGGCCCAACTGGTTCGCGGGCGGACCCCTTCGGGGGTC 2279043  
Qy 525 ThrThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGly 544  
Db 2279044 AGCGGGGTGGGACGCTCGCGCGCGCTCCGACGCGAGTCTGCTGTGTGGAGCCGCGG 2279103  
Qy 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
Db 2279104 GGCACCGGG-----GTCCAGTACACGACCGGGTTCGGCACCTCTGTGTGGGTCCGG 2279157  
Qy 565 GlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAla 584  
Db 2279158 GGCCTCCCGCGCGGGCGATCGTCGAGTCCGACCGGTCGACCGGACCTTCTACGGC 2279217  
Qy 585 LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAla 604  
Db 2279218 TTCAAGTCCGCGAGGTTCTAGCTCAGTTCGACGCGCGGGGACCTTCAACGGGTCCGG 2279277  
Qy 605 Ala---GlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys 623  
Db 2279278 GCCACGGGCTTCGCGAGCGGCGACAGC-----GTGCGCTTCAAGCGCTGCCCGGACG 2279331  
Qy 624 GluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHisSer 638  
Db 2279332 AAGGGGACATCTGGCTGGCGGCGGCGGACGCGGCGGTACGGGCTGTGGCACTCG 2279391  
Qy 639 ThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 2279392 ACGGACGCGCGCGGCGCTTCAACAAAGCTCCGACCGTCGACCGGCGGACACCATCGC 2279451  
Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly 678  
Db 2279452 TTCGCAAGCGCGGCGGCGGCGCTCGTACGAGAGCTTACACCGCGCGGAGATCGC 2279511  
Qy 679 GlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAsp 698  
Db 2279512 GGTGTGCGCGCATCTTCGGGTGACCGCAAGGCGCGGAGCTGGACCGCGGTCAACGAC 2279571  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2279572 GATCCACACAGTGGGTGGACGCGCGCGGATCACCGGTGACCCCGGCTTACGGG 2279631  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 2279632 CGCGTGTATGTGTGCGACGAGCGGCGGATCGTCTACGCGGACACCGCGGCTTCTCG 2279691  
Qy 739 SerGly 740  
Db 2279692 GACGGC 2279697

## RESULT 5

US-10-026-994-4  
; Sequence 4, Application US/10026994  
; Publication No. US20030113732A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Goedegebuur, Frits  
; APPLICANT: Ward, Michael

APPLICANT: Yao, Jian  
TITLE OF INVENTION: ECVI Endoglycanase and Nucleic Acids  
FILE REFERENCE: G0698  
CURRENT FILING DATE: 2002-04-30  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 2517  
TYPE: DNA  
ORGANISM: Trichoderma reesei  
US-10-026-994-4

Alignment Scores:  
Pred. No.: 7,51e-142 Length: 2517  
Score: 1579.00 Matches: 325  
Percent Similarity: 59.3% Conservative: 123  
Best Local Similarity: 43.0% Mismatches: 264  
Query Match: 39.1% Indels: 44  
DB: 5 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-4 (1-2517)

|    |     |  |     |
|----|-----|--|-----|
| QY | 6   | TyrThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPhe | 24  |
| DB | 61  | TTTTTCGGAAGAGCTCAAGCTCGCGCGCGCGCGCTTCGTCGCCGCGCATCATCTTC     | 120 |
| QY | 25  | AsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrp    | 44  |
| DB | 121 | CATCCCAAGACAAAGGCGTAGCATATGCAACAGATATTTGGCGGCTGTACCGCCTC     | 180 |
| QY | 45  | AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp        | 60  |
| DB | 181 | AAC---GCCGACGACTCATGACCGCGCTCAAGATGGATGCTGTAATGCCGCTGG       | 237 |
| QY | 61  | AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys | 80  |
| DB | 238 | CACAACTGG-----GGCATCGACGCTGTGGCTTGTATCGCGAGCATCAAAAG         | 288 |
| QY | 81  | ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu | 100 |
| DB | 289 | GTGTATCGCGAGTCGGCATGTATACGAACAGCTGGGATCGGATTAATGGAGCCATCAT   | 348 |
| QY | 101 | ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGly | 120 |
| DB | 349 | CGCTCGTCAGCCGCGCGCAACGTGGCTTCAACACTTGCCTTCAAAAGTCGGGGT       | 408 |
| QY | 121 | AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle | 140 |
| DB | 409 | AACATGCCAGACGCGGAGCGGAGAGCGTCTGGCTGTCTGATCGGCCCACTCCACATC    | 468 |
| QY | 141 | LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr | 160 |
| DB | 469 | ATCTACTTTGGTGTCTCGCTCAGGAAACGCGCTCTGGAAGTCTACGGACGGCGGTGACC  | 528 |
| QY | 161 | TrpSerGlnMetThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThr    | 180 |
| DB | 529 | TTTTTCAAGGCTCGTCTGCTCACGGCACTCGGACGTATACATCCAGACCCGAGTATTC   | 588 |
| QY | 181 | ThrGlyTyrGlnSerAspIleGlnGlyValTrpValAlaPheAspLysSerSer       | 200 |
| DB | 589 | AACGGCTACACACGACCAACAGGACTCATGTGGTTAGTTGCTACTCAACACGACG      | 648 |
| QY | 201 | SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro | 219 |
| DB | 649 | ACGACCGGGGAGCCACGCTCGTATCTTTGTGGCAGGCTGATAACATCACTGCTTCA     | 708 |
| QY | 220 | ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly | 239 |
| DB | 709 | GTCTATGTGAGCACGAATGCCGGCTCCACGTTGAGTGTGTACCGGGCAGCCAGGAAA    | 768 |
| QY | 240 | PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer | 259 |

|    |      |  |      |
|----|------|--|------|
| DB | 769  | TACTTTCTCAGAGCGGAAACTGCGAGCCAGAGAGAGCCCTTGATCTGACCTATTCC     | 828  |
| QY | 260  | AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer | 279  |
| DB | 829  | GATGCACAGGCGCGTATGATGCGACACTTGGCTAGTGTGGAGGTACGACATTCGAGGG   | 888  |
| QY | 280  | GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly | 299  |
| DB | 889  | GGAACCTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTA-----TACTTTGGC    | 939  |
| QY | 300  | TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle | 319  |
| DB | 940  | TTTGGCGGCTTGGCTCGATTGCAAAAGCCAGGAAACCTTGTGTGCTTCTTTGGAAC     | 999  |
| QY | 320  | SerTrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArg    | 339  |
| DB | 1000 | TCTTGGTGCCAGATGCTCAGCTGTTTGGTTCGACCGACTCTGGGACACATGAGGCGCG   | 1059 |
| QY | 340  | IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla | 359  |
| DB | 1060 | ATCTGGGCTGGCGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAACTCCCAAA   | 1119 |
| QY | 360  | GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer-----     | 375  |
| DB | 1120 | GCACCGTGGATCAAGAACAACTTTATTCGATGTGACGAGCGAGTCACCGTCGGATGCTC  | 1179 |
| QY | 376  | ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg | 394  |
| DB | 1180 | ATCAAGCGGCTCGGCTGGATGATTGAGTCTCTCGAGATGACCAACCCAGCAGCAACAC   | 1239 |
| QY | 395  | MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer | 414  |
| DB | 1240 | TGGCTCTACGCGACCGGATGACATCTTTGGCGGCCAGCATCTCACCACCTGGGACAG    | 1299 |
| QY | 415  | GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp    | 434  |
| DB | 1300 | CGCCCAATGTGTCAATCAATCACTGCGCAGCGGATCGAGGATTCCTCCGTCAGGAC     | 1359 |
| QY | 435  | LeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyPhe    | 454  |
| DB | 1360 | CTGGCTCTCGACCCCGGGAAGCGAGCTATTGGCGCGACTCGGAGCAGCAACCGGCTC    | 1419 |
| QY | 455  | ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal     | 471  |
| DB | 1420 | ACCTTTGCGACGAGAAACGACTCGGGACATCGCCGACAGCGTCTGGGCAAGCCCA      | 1479 |
| QY | 472  | PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla | 491  |
| DB | 1480 | TGGGCCACTCGACGAGCGTCTGACTACGCGGGAACCTCGGTCAAGAGCGTCTCGCGTC   | 1539 |
| QY | 492  | GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly | 511  |
| DB | 1540 | GCGAACACCGCGGACCAACAG-----GTGGCCATCTCGTCCGCGCGC              | 1584 |
| QY | 512  | GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla | 531  |
| DB | 1585 | GGCGCGAGTGGAGCATCGACTACGCGCGCGACACGCTCATGAACCGCGGCGCGGTGCC   | 1644 |
| QY | 532  | AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal | 551  |
| DB | 1645 | TATTCGGCGAGCGGACGACGATCTCTGTGTCAGCCGCTCTGTCGCGC-----GTG      | 1695 |
| QY | 552  | TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln | 571  |
| DB | 1696 | CAGCGCTCGAGTTCCAGGGCAGCTTGGCTCGCTCGAGCGTCCCGCGCGCGCGCTC      | 1755 |
| QY | 572  | IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr | 591  |
| DB | 1756 | ATCGCTCGGACCAAGAACCAACAGCGCTTCTTACCGCGGCTCCGGATCGACCTTTAC    | 1815 |
| QY | 592  | ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerGly    | 611  |



```

Db 1816 GTGACGACGACACCGCGCAGCTTC-----ACGCGGG-----CCCAAGCTGGC 1863
Qy
Qy 612 AlavalGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu 629
Db 1864 AGCGAGGAGCAGTCCGGATATCGTCTCACCCGACACCGCGGGCAGTGTGTATGTC 1923
Qy 630 AlaAlaSerSerGlyLeuTyHisSerThrAsnGlyGlySerSerTrpSerAlaIle--- 648
Db 1924 TCACACGACGTCGCGCATATTCCGCTCCACAGACTCGGGCAGCAGCTTTGGCCAAAGTCTCC 1983
Qy 649 ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrp 668
Db 1984 ACCGCTGACCAACACTACAGATCGCCCTGGGTGGGCTCA---GGCTCGAAGTGG 2040
Qy 669 ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer 686
Db 2041 ---AACCTGATCGCTTCGCGACC-----GGCCCGTCAGGGGCTCGCCTCTACGCCAGT 2091
Qy 687 AspAspCysGlyThrTrpTrpValLeuIleAsnAspAspGlnHisGlnTrpGlyAsnTrp 706
Db 2092 GGAGACAGCGCGCGCTCTCGTGGAGGACATCCAGGGCTCCAGGGCTTCGGCTCCATCGAC 2151
Qy 707 GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly 726
Db 2152 AGCACCAAGTCGCGGACGCGGAGCAGCCGCGGCGCAAGTCTACGTGGGCAACACGCG 2211
Qy 727 ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly 740
Db 2212 CGGGGCGCTTTTACGCTCAGGAGNACCGTCGGCGGCGGACGGCGGG 2259

RESULT 6
US-10-026-994-1
; Sequence 1, Application US/10026994
; Publication No. US20030113732A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Goedegebuur, Frits
; APPLICANT: Ward, Michael
; APPLICANT: Yao, Jian
; TITLE OF INVENTION: EGVI Endoglucanase and Nucleic Acids
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: GC698
; CURRENT APPLICATION NUMBER: US/10/026,994
; CURRENT FILING DATE: 2002-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2710
; TYPE: DNA
; ORGANISM: Trichoderma reesei
US-10-026-994-1

Alignment Scores:
Pred. No.: 8,26e-142 Length: 2710
Score: 1579.00 Matches: 325
Percent Similarity: 59.3% Conservative: 123
Best Local Similarity: 43.0% Mismatches: 264
Query Match: 39.1% Indels: 44
DB: 5 Gaps: 20

US-09-917-376-3 (1-740) x US-10-026-994-1 (1-2710)

Qy 6 TyrThrTrpSerAsnValAlaIle---GlyGlyGlyPheValAspGlyIleValPhe 24
Db 101 TTTTCATGAAGAACGTCAGCTCGCGCGCGCGCGGCTTCGTCGCCGATCATCTTC 160
Qy 25 AsnGluGlyAlaProGlyIleLeuTyValArgThrAspIleGlyGlyMetTyrArgTrp 44
Db 161 CATCCCAAGCAAAAGCGGTAGCATATGCACAGACATATTGGCGGCTGTACCGCCTC 220
Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp 60
Db 221 AAC--GCCGACGACTCATGACCGCGCTCACCGATGGGATTCTGTATGATCGCGCTGG 277

```

```

Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 278 CACAACATGG-----GGCATCGACGCTGTGGCTGTGATCCGACGACGATCAAAAG 328
Qy 81 ValTrpAlaAlaValAlGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 329 GTGTATCCCGCAGTCGGCATGTATACGAACAGCTGGGATCCGAGTAATAGGCCATCAT 388
Qy 101 ArgSerSerAspGlnGlyAlaValThrTrpGlnIleThrProLeuProPheLysLeuGly 120
Db 389 CGTCTGTCAGACCGCGCGGCGCAACGTGGCTCTCCAACTTGGCCCTTAAAGTCGGGGT 448
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
Db 449 AACATGCCAGGACGCGGAGCGGAGAGCGTCTGGCTGTGATCCGCGCAACTCCAAATC 508
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
Db 509 ATCTACTTTGGTGTCTCGCTCAGGAAACGCGCTCTGGAAGTCTACGAGCGCGCGTACC 568
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 569 TTTTCCAAAGGCTCTCGTTCACGGCAACTGGAGCTACATCCACAGCCCGAGTATTCC 628
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
Db 629 AACGGCTCAACACGACGACAGCAAGACTCATGTGGGTTCAGTTCGATCAACACGACG 688
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro---AsnAsnPro 219
Db 689 ACGACCGGGGAGCGACGCTCTGTTTGGCAGCGCTGATACATCAATCATCTGCTTCA 748
Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239
Db 749 GTCTATGTGACACGAATGCCGCTCCACGTGGAGTGTCTACCGCGGCGACCGAGGAAA 808
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSer 259
Db 809 TACTTTCTTCAAGGGCGAAACTGCGACGACAGAGAGGCTGTATCTGACCTATTTC 868
Qy 260 AsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279
Db 869 GATGGCACAGGGCGGTATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTTCAGGG 928
Qy 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGly 299
Db 929 GGAACCTTGGAAAGACATCACCCCTGTCTCTGGATCAGATCTA-----TACTTTGGC 979
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319
Db 980 TTTGGCGGCTTTGGCCTCGATTTCGAAAGCCGAGAACCTTGTGTGTCTTCTTTGAAC 1039
Qy 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGlyValaThrTrpThrArg 339
Db 1040 TCTTGTGGCCAGATGCTCAGCTGTTCGTCGACCGACTCTGGGACACACATGGAGCCCG 1099
Qy 340 IleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAla 359
Db 1100 ATCTGGCGTGGCGGAGCTATCCGACTGAGACCTATTACTACAGCATCTCAATCCCAAA 1159
Qy 360 GluProTrpLeuThrPheGly---ValGlnProAsnProProValProSer----- 375
Db 1160 GCACCGTGGATCAAGAACAACTTTATCGATGTGACGAGCGAGTCACTCCGATGGTCTC 1219
Qy 376 ---ProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArg 394
Db 1220 ATCAGCGCTCGCTGATGATTGAGTCTCTCGAGATTGACCAACCGACGACGACACAC 1279
Qy 395 MetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSer 414
Db 1280 TGGCTCTACGCGACCGGAATGACAATCTTTTGGCGGCCACGATCTCAACCACTGGGACG 1339

```



|    |      |  |      |
|----|------|--|------|
| Qy | 415  | GlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAsp      | 434  |
| Db | 1340 | CGGCACAAATGTCAATCCAAATCGCATCGCAGACGGATTCGAGAAATTCCTCGGTCCAGGAC | 1399 |
| Qy | 435  | LeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPhe      | 454  |
| Db | 1400 | CTGGCTCTGCACCCGGCGAAGAGGAGCTATTGGCCGAGTCGGAGACGACACACGGCTTC    | 1459 |
| Qy | 455  | ThrHisAla-----AspValThrAlaValProSerThrIlePheThrSerProVal       | 471  |
| Db | 1460 | ACCTTTGCCAGCAGAAACGACCTCGGACATCGCCGACAGCGTCTGGGCAACGCCACCA     | 1519 |
| Qy | 472  | PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla   | 491  |
| Db | 1520 | TGGGCCACCTCGACGAGCGTCGACTACGCCGGGAACCTCGGTCAAGACGCTGTCGCGGCT   | 1579 |
| Qy | 492  | GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly   | 511  |
| Db | 1580 | GGCAACACCCCGCGCAGCAACAG-----GTGGCCATCTCTGTCGACGGC              | 1624 |
| Qy | 512  | GlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAla   | 531  |
| Db | 1625 | GGCGCAGCTGGAGCATCGACTACGGCGCCGACACAGTCCATGAACGGCGCACGGTGGCC    | 1684 |
| Qy | 532  | AlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValVal   | 551  |
| Db | 1685 | TATTTCGGCCACCGCGCAGACAGCATCTCTGTGCTCAGCCGCTCGTCCGCGC-----GTG   | 1735 |
| Qy | 552  | TyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGln   | 571  |
| Db | 1736 | CAGCGCTCGCAGTTCAGGGCAGCTTTGCTCGCTCTCGAGCTGCGCGGGCGCGCGCT       | 1795 |
| Qy | 572  | IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyr   | 591  |
| Db | 1796 | ATCGCTTCGACAGAGAACCAACAGCGCTCTTCTAAGCCGGCTCCGATCGACCTTTTAC     | 1855 |
| Qy | 592  | ArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGly   | 611  |
| Db | 1856 | GTACGACGACACCGCAGCAGCTTC-----ACGCGCGGG-----CCCAAGCTGGGC        | 1903 |
| Qy | 612  | AlaValGlyValMet-----PheHisAlaValProGlyLysGluGlyAspLeuTrpLeu    | 629  |
| Db | 1904 | AGCGAGGACGATCCGGATATCGCTGTCTCACCCGACACCGCGGGCAGCTGTGTATGTC     | 1963 |
| Qy | 630  | AlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIle---   | 648  |
| Db | 1964 | TCGACCGACGTCGGCATATTCCGCTCCACAGACTCGGGCACGACCTTTGGCCAAAGTCTCC  | 2023 |
| Qy | 649  | ThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlyLysSerTyr   | 668  |
| Db | 2024 | ACGCGCTTACCAACACCTACAGATCGCCCTGGTGTGGGCTCA---GGCTCGAACTGG      | 2080 |
| Qy | 669  | ProAlaValPheValValGlyThrIleGlyGlyValThrGlyAla-----TyrArgSer    | 686  |
| Db | 2081 | ---AACCTGTATGCTTCGGCACCC-----GGCCCGTCAGGGGCTCGCTCTACGCCAGT     | 2131 |
| Qy | 687  | AspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrp   | 706  |
| Db | 2132 | GGAGACAGCGGCGCCCTCTGGACGGACATCCAGGGCTCCACGGGCTTCGCGCTCCATCGAC  | 2191 |
| Qy | 707  | GlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGly      | 726  |
| Db | 2192 | AGCACCAAGTCCCGCGCAGCGCAGCACCGCCGGGCAGTCTACGTGGGCACCAACGGC      | 2251 |
| Qy | 727  | ArgGlyIleValTyr-----GlyAspIleGlyGlyAlaProSerGly                | 740  |
| Db | 2252 | CGGGCGCTCTTTTACGCTCAGGAACCGCTCGCGGGCGCAGCGGGCGG                | 2299 |

Db 622 TCCGGCAGGAGTCGTGTTCC-----CTCGTCGCCGCC 654  
Qy 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val 220  
Db 655 GGGCGCACCGTCTACCGCGCGTGGGTGACGCGACGCGCCTCGGCGACGCGAACCTG 714  
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240  
Db 715 TACCGCACGCGCGAC--GGCACACCTTGGGGGCGGTCCCGCGCGCGCTCCGCGCAC 771  
Qy 241 -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
Db 772 TCCGCCAGGTCCCGTCCGCGCGCGGTACACGACGCGCACCCCGGAGCTGTAGTGACG 831  
Qy 258 ThrSerAsnThrGlyGlyProTyrAspGlySerGlyAspValTrpLysPheSerVal 277  
Db 832 TACGGCAGCAGCACCGCGCGCGCGCGCGTCCGACGCGCGGTGCAACAGTGGTAC 891  
Qy 278 ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr 293  
Db 892 GCCACGGGAGCTGACCGAGGTCAACCGGTGAAGCGCGCGGAGCAGCAGGCGACGCG 951  
Qy 294 AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313  
Db 952 TCGCGGACACCTTCGCTACGCGGGGTGCGGTGCGCGCGCGCGCGCGCGCACCTC 1011  
Qy 314 MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly 333  
Db 1012 GTGCTCTCCACCAACACCGCTGGCGCGACGCGCGCGGTCTTCGGTCCACGCGCGC 1071  
Qy 334 GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr 353  
Db 1072 GCGCGTACCTGGAGCTCCCTCAAGGAC-----GCCGCC 1104  
Qy 354 ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProVal 373  
Db 1105 GTGTTCAGCGTCCGAGACTCCCTCTCAGCTGGGCGCGAC----- 1149  
Qy 374 ProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIleAspProPheAsnSerAsp 393  
Db 1150 ---NAGCGAGTTCGGCTGGTGTATCCAGCGGTCCGCGGTGCGACCCGCTACGATCCAG 1206  
Qy 394 ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp 413  
Db 1207 CACGTCTGTACGGACCGCGCGGACCTCTACGCGCACCGCGCGACCTCAAGCGTGG-- 1263  
Qy 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
Db 1264 -----GCACCGCGGATCCGCGCGCTGGAGGAGAGCGCGGTGGC 1302  
Qy 434 AspLeuIleSerProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly 452  
Db 1303 CAACGTATCTGCCCGCGTCCGGAGGACACCTGTACGCGACTCGGGGACATCGGT 1362  
Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThr-----IlePheTrpSerProVal 471  
Db 1363 GTGATGTACACGAGCGCTCACGCGGTCTCGTCCGCGCGCATGGCGACGACCCCGTG 1422  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1423 TTCGGTTCGGGACGCGGACCTCGCGAGCGCGCGCGCGGTATGCTCGCGCACG 1482  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1483 GGCTGGGCGCACCGCACCGCACCGC-----GCGTACTCCACGACGCGC 1524  
Qy 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr 529  
Db 1525 GGGGGGACCTGGGCGCGCTTCAGCGCCCGCGCGACATCGCCACGAGCGCACCGGACCG 1584  
Qy 530 ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp 545  
Db 1585 ATCGCCACAGTGCAGCGCGCGCGCACACTGTGTGCTCTCTGCTGCTGCTGCGACGCGC 1641

Qy 546 ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla 561  
Db 1642 -----ACGACGTACGCGCGCCACCGCTCGACGCAACGCGCGAGCTGGTCC 1689  
Qy 562 AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr 581  
Db 1690 GAGGTCTCTCTCCCGAAGGCGCGCACACCGGTGCGCGACCGCGCGCTACCGCGCGC 1749  
Qy 582 PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThr 599  
Db 1750 TTCTACGCTACGACTTCGACATGGAACGCTATACCGCAGCAGCTGACAGTGGCCGCTCG 1809  
Qy 600 PheGlnProValAlaAlaGlyLeuProSerSerGlyValAlaValGlyValMetPheHisAla 619  
Db 1810 TTCACGCGCGTTCGCGCGGCGACTGCGC---TCCGCGCAGACGCGAGTCAAGTGGTGGCG 1866  
Qy 620 ValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer 638  
Db 1867 GCGCGCGGACGAAAGCGCGGACCTTGTGGCTCTCCGCAATGGAACGGGCTCTACCGGTCC 1926  
Qy 639 ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly 658  
Db 1927 ACCGACGCGGAGCACCTTCGCGAGTTCGACAGTCTGGGCTCTGACACCTTCGCGC 1986  
Qy 659 PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValValGlyThrIleGly 678  
Db 1987 TTCGCAAGCGCGCGCGCGCGGCGGCGGACTACCGCGGATCTACGAGTCCGCTCGCGGAG 2046  
Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698  
Db 2047 ACCATCACCGCGTCTACCGCTCCGACGCGCGCGGACGACATGGGTCCGGATCAACGAC 2106  
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718  
Db 2107 GACGCCACCGAGTGGGTGGATCGCGCGGCGCGGTCTGCTGCGTACCCCGCGCATCCACGCG 2166  
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733  
Db 2167 CGGTCTACCTCGCCACCAACGCGCGCGCGCATCCAGTACGCGGAG 2211

## RESULT 8

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1  
Alignment Scores:

|   |           |  |         |
|---|-----------|--|---------|
| Pred. No.:  | 1.61e-120 | Length:  | 9025608 |
| Score:  | 1405.00   | Matches:   | 308     |
| Percent Similarity:                                   | 54.6%     | Conservative:  | 104     |
| Best Local Similarity:                                | 40.8%     | Mismatches:  | 257     |
| Query Match:  | 34.8%     | Indels:  | 86      |
| DB:   | 6         | Gaps:  | 21      |
| US-09-917-376-3 (1-740) x US-10-156-761-1 (1-9025608) |           |  |         |
| Qy  | 6         | TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn    | 25      |
| Db  | 3161882   | TACCGTGGCGCAACCGCGTTCATCGGGGACCGGCTTCGTACCGCGGTCTCTTCCAC     | 3161823 |
| Qy  | 26        | GlulGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp   | 45      |
| Db  | 3161822   | CCCTCCGTACCGCGTCTCGCCTACGCCCGGACACATCGCGCGCGCTACCGTGGGAC     | 3161763 |
| Qy  | 46        | AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr | 65      |
| Db  | 3161762   | GACCGCGCGCGTGGACCCCGCTCATCGACACCTCGGCTGGGACGACTGGGAACCTC     | 3161703 |
| Qy  | 66        | AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal | 85      |
| Db  | 3161702   | CTCGCGTTCGAGGCGATGCGCTCGACCCGACCGACCGCGGCTCTACTCGCCGTG       | 3161643 |
| Qy  | 86        | GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerAspGln    | 105     |
| Db  | 3161642   | GGCACCTACGCCCGAGTCGTGGCGGCAAC--GGCGGCGTCTCGCTCCGAGGACCGC     | 3161586 |
| Qy  | 106       | GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg | 125     |
| Db  | 3161585   | GCGCCACCTCGGACCGCGCACCGCTGACCGTGAAGCTCGCGGCAACGAGGACGCGCG    | 3161526 |
| Qy  | 126       | GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla | 145     |
| Db  | 3161525   | GCGCGGTGAGGACTCTCTGACCCGCGGACGCGACACCCCTCTGGTGGGACG          | 3161466 |
| Qy  | 146       | ProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr | 165     |
| Db  | 3161465   | --CGGCACGACGGGTGCTCAAGTCGACCGCGGCGGCGCACTTGGGACCGCGGAC       | 3161409 |
| Qy  | 166       | AsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGlnSer | 185     |
| Db  | 3161408   | GCCTTCCCG-----GCGAAGCGCAACTCC                                | 3161385 |
| Qy  | 186       | AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGlnAla | 205     |
| Db  | 3161384   | TCCGGGACGGGAGTCGTGTTCTC-----CTCGTCGCGCGCC                    | 3161352 |
| Qy  | 206       | SerLysThrIlePheValGlyValAlaAspProAsnAsnPro-----Val           | 220     |
| Db  | 3161351   | GGGCGCACCGTCTAGCGCGGTGGGTGACGGGACGCGACCTCGGGACGCGGACCTG      | 3161292 |
| Qy  | 221       | PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe | 240     |
| Db  | 3161291   | TACCGCACGGCGGAC--GGCACGACCTGGGGGCGGTCCCGCGCGCCCTCCGCGCAC     | 3161235 |
| Qy  | 241       | -----IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla     | 257     |
| Db  | 3161234   | TCCGCCAAGGTCCCGCTCGCGCGCGGTACGACACGACACCCCGCGAGCTGTAGCTGAC   | 3161175 |
| Qy  | 258       | ThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerVal | 277     |
| Db  | 3161174   | TACGGCAGCGACCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG      | 3161115 |
| Qy  | 278       | ThrSerGlyThrTrpThrArgIleSerProVal-----ProSerThrAspThr        | 293     |
| Db  | 3161114   | GCCAGGGGACGTGACCGAGGTACCCCGGTGAAGCGGCGGCGGACGACGAGCGACGCG    | 3161055 |
| Qy  | 294       | AlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle | 313     |
| Db  | 3161054   | TCGGCGGACACTTCGCTACGGCGGGTCTCGGCTCGACGCGCGCGCGCGCGCGCGCGC    | 3160995 |
| Qy  | 314       | MetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGly | 333     |
| Db  | 3160994   | GTCTGCTCCCAACCAACACCGTGGCGGCGGACACCGGTCTCTCGGTCCACGAGCGGC    | 3160935 |
| Qy  | 334       | GlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeuArgTyr | 353     |
| Db  | 3160934   | GGCGGTACTGGAGCTCCCTCAAGGAC-----GCCGCC                        | 3160902 |
| Qy  | 354       | ValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnProProVal | 373     |
| Db  | 3160901   | GTGTTTCGAGCTGTCGAGACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT    | 3160857 |
| Qy  | 374       | ProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspPropheAsnSerAsp | 393     |
| Db  | 3160856   | ---AAGCCGGAAGTTCGGCTGGTGGATCCAGGCGCTCCGCGTCCAGCCGTACGACTCC   | 3160800 |
| Qy  | 394       | ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp | 413     |
| Db  | 3160799   | CACGTGCTGTCGAGACCGCGCGGACCTCTTACGGCACCGCGACCTCAAGCGCTGG--    | 3160743 |
| Qy  | 414       | SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn    | 433     |
| Db  | 3160742   | -----GCACCGCGATCCCGCGCTGGAGGAGCGCGCTGCGC                     | 3160704 |
| Qy  | 434       | AspLeuIleSerProProSerGly---AlaProLeuIleSerAlaLeuGlyAspLeuGly | 452     |
| Db  | 3160703   | CAACTGATCTCGCCCGCTCGGGGAGGACACCTGATCAGCGGACTCGGGGACATCGGT    | 3160644 |
| Qy  | 453       | GlyPheThrHisAlaAspValThrAlaValProSerThr---IlePheThrSerProVal | 471     |
| Db  | 3160643   | GTGATGTACACGAGCGGTCTACGGCGTCTCGTTCGCGCGCATCGGACGACCCCGTG     | 3160584 |
| Qy  | 472       | PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla | 491     |
| Db  | 3160583   | TTCGGTGGCGACGAGACTCGCGGCGCGCGCGCGCGCGCGGTATGTCTCGCGACG       | 3160524 |
| Qy  | 492       | GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly | 511     |
| Db  | 3160523   | GGCTGGGCGCACCGGCAACGCG-----GCGTACTCCACGACGCGC                | 3160482 |
| Qy  | 512       | GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThrGlyGlyThr  | 529     |
| Db  | 3160481   | GGCGGACGCTGGCGCGCTTCGAGGCGCCACGCGCATCGCCAAAGGACGCGCGGACCG    | 3160422 |
| Qy  | 530       | ValAlaAlaSerAlaAspGlySerArgPheValTrpAla-----ProGlyAsp        | 545     |
| Db  | 3160421   | ATCGCCACGAGTCCGCGCGCGCGGACACTGTGTGTCTCTCTCTCTCTCTCTCTCTCT    | 3160365 |
| Qy  | 546       | ProGlyGlnProValValTyrAla-----ValGlyPheGlyAsnSerTrpAla        | 561     |
| Db  | 3160364   | -----ACAGCGTACCGGCGCCACCGCTCGACGAGCAACCGCGGAGCTGGTCC         | 3160317 |
| Qy  | 562       | AlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThr | 581     |
| Db  | 3160316   | GAGGTCTCTCTCTCCGAAAGCGCGCACACCGTTCGCGCGCGCGCGCGCGCGCGCGCG    | 3160257 |
| Qy  | 582       | PheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyValThr     | 599     |
| Db  | 3160256   | TTCTACGCTTACGACTTCGACAACTATACGCGGACGCTATACGCGGACGACTGAGTGG   | 3160197 |
| Qy  | 600       | PheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAla | 619     |
| Db  | 3160196   | TTACGCGCGGTGGCGGAGTCTGCCC-----TCGGCGACGCGGAGTTCAGCTGGTGGC    | 3160140 |
| Qy  | 620       | ValProGlyLysGlyAspLeuTrpLeuAlaAlaSer---SerGlyLeuTyrHisSer    | 638     |
| Db  | 3160139   | GCGCGGAGCGGAGCGGACCTGTGGTCTTCGCCAAATGGAACGGGCTCTACCGGTCC     | 3160080 |
| Qy  | 639       | ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly | 658     |
| Db  | 3160079   | ACCGACGCGGCGGACACCTTCGCCAGGATCGACAGCTGTGGGCTCTGTACACCTCGGC   | 3160020 |
| Qy  | 659       | PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGly    | 678     |

```
Db 3160019 TTCCGCAAGCGCGCGCGCGCGCGCTACCCGGCGATCTACCAAGGTGGCTCGACGGAG 3159960
Qy 679 GlyValThrGlyAlaTyrArgSerAspCysGlyThrThrTrpValLeuIleAsnAsp 698
Db 3159959 ACCATACCGCCCGCTACCGCTCGACGACCGCCAGGACATGGTTCGGATCAACGAC 3159900
Qy 699 AspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArg 718
Db 3159899 GAGCCCAACAGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3159840
Qy 719 ArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGlyAsp 733
Db 3159839 CCGGTCTACCTCGCCACCAACGGCGCGCGCATCCAGTACGCGGGAG 3159795

RESULT 9
US-09-927-827-21
; Sequence 21, Application US/09927827
; Publication No. US20030036178A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 21
; LENGTH: 3668
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)...(2668)
US-09-927-827-21

Alignment Scores:
Pred. No.: 3,77e-99 Length: 3668
Score: 1139.00 Matches: 267
Percent Similarity: 50.9% Conservative: 120
Best Local Similarity: 35.1% Mismatches: 295
Query Match: 28.2% Indels: 79
DB: 21 Gaps: 21

US-09-917-376-3 (1-740) x US-09-927-827-21 (1-3668)
Qy 1 AlaThrThrGlnProTyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAsp 20
Db 1217 GCCACGTCGCGGCGCTACCGAGTGGCGCAGTGTGCCATTTGGCGTGGCGCTTTGTCA 1276
Qy 21 GlyIleValPheAsnGluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGly 40
Db 1277 GGTGTGTGTTCATCCCGCGCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1336
Qy 41 MetTyrArgTrpAspAlaAlaAsnGlyArgTyrIleProIleuLeuAspTrpValGlyTrp 60
Db 1337 GCGTACCGCTGGGATGGCGAGCGACGAGTGGACCGCGCTGACCGAGTGTGTGGCGCT 1396
Qy 61 AsnAsnTrpGlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLys 80
Db 1397 GACGACTGGAACTGATGGGATCGACGATTCGCCGTGACCCCGCGGATGCCGATGGC 1456
Qy 81 ValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeu 100
Db 1457 CTGTATCTGGCGCGCGCACCTATATCATGAA---CGCGCGCGCATCTGCCGAGTGTG 1513
Qy 101 ArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProIleuLeuAspPheLeuGlyGly 120
Db 1514 CGTCTGTCAACCGCGCGCGCGCGCTTCGAGCGTCCGACCTGCGGTAAAGCTGGTGTGT 1573
Qy 121 AsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIle 140
```

```
Db 1574 AACCACTGTGGCGCGCGCAATGGCGAGCGGTGGCGGTGGACCGCACATGGCCCGCGTG 1633
Qy 141 LeuTyrPheGlyAlaProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThr 160
Db 1634 CTGCTGTGGCTCGCGGATGCC---GGCTGTGGCTAGGACGATCGCGCGCGCGAC 1690
Qy 161 TrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThr 180
Db 1691 TGGCGAAGTGGCGTCTTTCCGACGCGCGCGTGGCGGTGGCCACCGCGCGCAATCAT 1750
Qy 181 ThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 200
Db 1751 GTTGGCGCGAGACGAGCGGTG---GGGATCGCTTTGTGTGTTCGACGACGACGATGGC 1807
Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProVal 220
Db 1808 AACACGGCTCGCCMACACCGCGCATCTACGTGGCGGTGTCCACCGAACAGACCGCGTG 1867
Qy 221 PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPhe 240
Db 1868 TATGTGTCCGAAGATGCGCGCGCAGTTGGGCACCGGTGGCGGCGCAACACCGCGCGCTG 1927
Qy 241 IleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAlaThrSerAsn 260
Db 1928 CGCCGAGCCACATGGCGCGCGCGCAGCGATGGGCAC---TGGTATCTGAGCTATGGCGAC 1984
Qy 261 ThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGly 280
Db 1985 CAGCCCGCGCGCGACCTGATCGCGGGGAGCCTTTGTGGAATTCACCGCGCGCACAGGG 2044
Qy 281 ThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTyrPheGlyTyr 300
Db 2045 CGTGGCGTGAGATCAGCCCGATTCGCGAG---CCAGCCAGTGGCGATGGATTCGCGTGG 2101
Qy 301 SerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSer 320
Db 2102 GGTGGGTGGCGGTGGATCCGCAACATCCGAGGTGTCTGTGGCGCAGCACCTTCGCGCGT 2161
Qy 321 TrpTrpProAspThrIlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIle 340
Db 2162 CGCAGCGCGCGCGCAGCAGCTATGTCGACGCGTGGTGGCAAGCAGCTGGAGCGCGTGG 2221
Qy 341 TrpAspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGlu 360
Db 2222 -----TTGGCGCGATGGGTGTTCATCCAGCGCGCGG 2254
Qy 361 ProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTyr 380
Db 2255 CCGTGG---ACCGCACATGCCACCGCGCAC-----TGG 2284
Qy 381 MetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThrGly 400
Db 2285 ATGGGG---GGCTGGCGCATCGATCCGTTTCAGCGCAACCATCGCTGTGTGACCGCG 2341
Qy 401 AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIle 420
Db 2342 TACGCGCATCTGGCGCTCGCGCAATCTG-----CAGGATTTTCGCC 2380
Qy 421 AlaPro-----MetValLysGlyLeuGluThrAlaVal 432
Db 2381 GCACCGCAGCGCGCTGCAGTGGTGTTCAGGACCGGTGGCTGGAGAAACCGTGGCG 2440
Qy 433 AsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGly 452
Db 2441 CTGACCTGTCTCAGCCCGATGGCTGGCGCGCATCTGCTCAGCGCGCTCGCGCATATCGAC 2500
Qy 453 GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPhe 472
Db 2501 GGCTTCGCGCATGACGACCTGGACCGCGTG---CAGTTGCGAGTACGCGCGCGCGCGGTG 2557
Qy 473 ThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGly 492
```

[illegible][illegible]

Db 655 ACGGCGCGTTCCGGCAAGACCGCGTGTGATCGCGCGGAGCCATGAAGTGCCT 714  
Qy 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
Db 715 CTCACCCCC-----AACTTCTCTACGTGATTCACGCCGACTACCTGGTTCATGGCG 768  
Qy 267 GlySerSerGlyAspValTyrLysPheSerValThrSerGlyThrTrpThrArgIle--- 285  
Db 769 GTACAGTTCGGCGAAGTCTGGCGCAAGACCGACCTCGCGCGCTTGGGACGACATTACT 828  
Qy 286 -----SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
Db 829 CCGCGCTCGCAACTCGTGCCTGCCCTGTACAAACACGAGAGTTCCTCGCGGCGGA 888  
Qy 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 889 TTTTGGGTCTCAGCGTCGAGCGGACCAACCCCAACCGTCTCGTGGTCAACACC---CTC 945  
Qy 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyGlyAlaThr 336  
Db 946 GACCGGACCCCGGACCGCGCTCGACAGCATCTACCTCTCAACCGATCGCGCGGAC 1005  
Qy 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
Db 1006 TGAAGACAGCTACCCAGCTCTGTCCCGTCCAACTCGAAGTAACCTGGGCGCACCCG 1065  
Qy 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
Db 1066 ACTAACCGCGCGGTAC---AAGACGCGACCGCTCTTCGGTGGTCTCGACTTCAACAC 1122  
Qy 368 GlnPro-----AsnProValProSerPro-----LysLeuGly 379  
Db 1123 GGTCCCGAGTGGGGGGGATACGGTGCAGCGGTACCGCGCGCTCAACCAAGTTGGC 1182  
Qy 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399  
Db 1183 TGTGGATGAGCGGTGTGCTTATCGATTCGTTCAACCCCGAGCACCTGTATGTCAGCG 1242  
Qy 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TrpAspSerGly 415  
Db 1243 GGGGCGACATCTGGGCGAGCGACACGCTCTCCGTGTCCAGAGGACTCG----- 1293  
Qy 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluThrAla 431  
Db 1294 -----GGCCGAGCTGTGTACTCCAGATCGACGATATCGAGGAGATGCG 1338  
Qy 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
Db 1339 ATCTGTGCTCGCTCGCCCAAGAGCGCGCGCTCTCTCGGCGCATCGGTGACATT 1398  
Qy 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
Db 1399 AGCGCATGAGACGACGACCTCACCAAG---CCCAAGAGATGTTGTGGCGCCCG 1455  
Qy 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
Db 1456 TTCTCCAACTCGACACGATCGAGCTCGCGGCACTTCCCAACGTTGTCGTCGCGGCC 1515  
Qy 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
Db 1516 GGATCTCTCGGACACGAGTACGACGCGCGTGGCGCGGCTGCTACGCGACTGACGCG 1575  
Qy 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThr----- 526  
Db 1576 GGAGACGCGTGGACCATCTTCTCCCTACTCTGCTCCCTGTGATGAACGCGACCTAC 1635  
Qy 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544  
Db 1636 GGCAGACGATGATGATGTCGAGCGGCGGACGATCGTGTGTGTCGACCAAGCTTGAC 1695  
Qy 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
Db 1696 GAGCAGGCTCGGACCGTGTACTCGCAGCACTATGCAAGACGTGTCT----- 1746

Qy 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578  
Db 1747 ---GTTCCCGTGGGACCTGAAGGCCCCAGACTGCAATGTCTCTCGGACAAGGTCCAG 1803  
Qy 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598  
Db 1804 GATGGCAGTTCTACGTACCGATGGCGGCAAGTTCTTCGTCGACCGACGCGGGAAG 1863  
Qy 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
Db 1864 TCGTATCCCGCAAGGGCGCGGACTGTGCTACT-----GGCACATCGCTCATGCTGCC 1917  
Qy 619 AlaValProGlyLysGluGlyAspLeuTrpLeu-----AlaAlaSerSerGlyLeuTyrHis 637  
Db 1918 GTGAACCCCTGGGTGGCGGCGAGCGTCTGGTGGCTGTTCGAGGGCGGTCTCTTCCAC 1977  
Qy 638 SerThrAsnGlyGlySerSerTrpSerAlaIle-----ThrGlyValSerSerAlaValAsn 656  
Db 1978 TCGACCGACTTGGCGCTCGTTTACGAGGGTGTGTACGAGGTAGTACCGCAACGCGAC 2037  
Qy 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
Db 2038 GTCCGCCCCCAAGTCAAGTCGACGCGGCAAGAGGTAGCGCCCTCCCGGTCTTC 2097  
Qy 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGly 690  
Db 2098 ATCTGGGCGACCGACAAGCCCTGGAAGCGACATCGCGCTGTACCGTCCGACGACAACGG 2157  
Qy 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
Db 2158 AGCACTGGAGCGCGCTCAATGACGAGGACGACCACTACTCGGGC---CCCACCATGATC 2214  
Qy 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
Db 2215 GAGCGCACCCCAAGGTCTACGCGCGGTGTATCTAGGACGACGCGCGGTATCGTG 2274  
Qy 731 TyrGlyAspIle 734  
Db 2275 TACCGCGACCTT 2286  
RESULT 11  
US-10-395-241-11  
; Sequence 11, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLYCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: 073756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 11  
; LENGTH: 2646  
; TYPE: DNA  
; ORGANISM: Geotrichum sp. M128  
; FEATURES:  
; NAME/KEY: CDS  
; LOCATION: (120)..(2558)  
US-10-395-241-11  
Alignment Scores:  
Pred. No.: 3,31e-97 Length: 2646  
Score: 1117.00 Matches: 281  
Percent Similarity: 50.4% Conservative: 124  
Best Local Similarity: 35.0% Mismatches: 279  
Query Match: 27.7% Indels: 120  
DB: 7 Gaps: 31

US-09-917-376-3 (1-740) x US-10-395-241-11 (1-2646)

QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyValPheAsn 25  
 DB 198 TACGAGTTCAAGATGTCGCGATCGCGCGCGGGTACATTACCGGATTGTTCGCGCAC 257  
 QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrpAsp 45  
 DB 258 CCAAGACCAAGGACCTGCTGTACGCGCGCACGACATTGCGCGCGCTACCGTGGGAC 317  
 QY 46 AlaAlaAsnGlyArgTyrIleProLeuLeuAspTrpValGlyTyrAsnAsnTrpGlyTyr 65  
 DB 318 GCAGCGCTCAAGTGGATCCCGCTCAACGACTTATCGAGCGCGACGATGAACATT 377  
 QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 DB 378 ATGGCGACGAGTCGATCGCGCTGGACCCCAACACCCGACGAGCTGTACTCGCGCAG 437  
 QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
 DB 438 GGGCGCTATGTCGCGCACGAGTGG---GCGCGGTTCTATGTCTCCGACAC 485  
 QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly 124  
 DB 486 CGCGCGCATGCTTTACAATCTACGAGTCGCGCTTCCGATGGCGGCCAACGACATGGGA 545  
 QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
 DB 546 CGCAACATGGCAGCGCTCGCTGTCAACCGTTCAACTCGAACGAGTCTGGATGGGT 605  
 QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
 DB 606 ACGGTACA---GAGGGTATCTGAAGATTCGACCGCGCCCAACACCTTGCACAAACGTC 662  
 QY 165 ThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
 DB 663 ACGTCCATCCCGACGCGGTTTC---ACCAACGGTATCGGATACACG 704  
 QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGlyGln 204  
 DB 705 TCG---GTCAATTTTCGACCC---GAA 725  
 QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
 DB 726 CGTAATGGCACCATCTACGCGAGCGGACTCCCGCGCAGGCG---ATGACTGCACGCAC 782  
 QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle--- 241  
 DB 783 GACGCGGTGTCTGTTGGGACCACTGCGCGCGCAGCGCTCCAGCTGGCTCAACAGGACC 842  
 QY 242 ---ProHisLysGlyVal 246  
 DB 843 ACGGCGCGTTCCCGGCAAGACCGCGTCGATCGCGCGCGACGCCCATGAAGTCGCT 902  
 QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
 DB 903 CTCACCC---AATTCCTCTAGTGAATACGCCGACTACCTGGTTCATATGGCG 956  
 QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgile--- 285  
 DB 957 GTCAGTTGCGGAGTCTGCGCGCCAGACCGACCTCGGCGGCTGGGACGACATTACT 1016  
 QY 286 ---SerProValProSerThrAspThrAlaAsnAspTyrPheGly 299  
 DB 1017 CCGCGCGTCGCGCAACTCGCTCGCTGCCGTACCAACACACGAGCTTCCCTCGCGCGGGA 1076  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 1077 TTTTGGCGTCTACGCGTCGACGCGACCAACCCCAACCGTCTCGTGTATCACC---CTC 1133  
 QY 320 SerTrpTrpProAspThrIle---IlePheArgSerThrAspGlyGlyAlaThr 336

DB 1134 GACCGGACCCCGGACCCCGCTCGACACGATCTACTCTCAACCGATGCCGCGCGACC 1193  
 QY 337 TrpThrArgIleTrpAspTrpThrSer---TyrPro 347  
 DB 1194 TGAAGAGCAGTCACCCAGCTCTCGTCCCGTCCAACCTCGAAGGTAACCTGGGCGCACCG 1253  
 QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
 DB 1254 ACTAACGCGCGCGGTAC---AAGGACGCGCACGCTGTTCCGTGGCTCGACTTCAACAAC 1310  
 QY 368 GlnPro---AsnProProValProSerPro---LysLeuGly 379  
 DB 1311 GGTCCCGAGTGGGGGGATACGTCGCGCGCACGCTACCGCCCGCTCACCAGTTTGGC 1370  
 QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399  
 DB 1371 TGGTGGATGAGCGCTGTATCGATCGTTCAACCCCGACGACCTGATGATGACGACG 1430  
 QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys---TrpAspSerGly 415  
 DB 1431 GGGCGCACCATCTGGCGCACCGACGCTCTCCGTGTGAGAGGACTGG--- 1481  
 QY 416 GlyGlnIleHisIleAlaPro---MetValLysGlyLeuGluGluThrAla 431  
 DB 1482 ---GCGCGAGCTGGTACTCCAGATCGACGCTATCGAGGAGATGCG 1526  
 QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
 DB 1527 ATCTGTGCTCGCTCGCGCAAGCGCGCGGCTCTGTCTGGGCTCGGTGACATT 1586  
 QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
 DB 1587 AGCGCATGAACGACGACGACCTCACCAG---CCCGACAGATGTTTGGTGGCCCGCAG 1643  
 QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
 DB 1644 TTCTCAACCTCGACAGCATCGACGTCGCGGCAACTTCCCAACGTTGTCTGCGCGCC 1703  
 QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
 DB 1704 GGATCCTCGGACACGAGTACGACGCGCTCGCGCGGCTGCGTACGCGATGACGCG 1763  
 QY 512 GlyLysAsnTrp---PheGlnGlySerGluProGlyGlyValThrThr--- 526  
 DB 1764 GGAGACGCTGGACCATCTTCTCTACCTCCCTCTCTGGCATGACGCGGACCTACCAG 1823  
 QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro---Gly 544  
 DB 1824 GGCAGCAGATGTCAGTCGACGCGAGCGCGCAGATCGTGTGTGTCGACCAAGCTTGAC 1883  
 QY 545 AspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
 DB 1884 GAGCAGGCTCGGACCCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1934  
 QY 565 GlyValProAla---AsnAlaGlnIleArgSerAspArgValAsn 578  
 DB 1935 ---GTTCCCGTGGCGACCTGAAGCCCGACCTGCAATGCTCTCGACAGGTCAG 1991  
 QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyVal 598  
 DB 1992 GATGCGACGTTCTACGCTACCGATCGCGCAAGTCTTCTGCTCGACCGCGCGGAG 2051  
 QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
 DB 2052 TCGTATGCGCGCAAGGCGCGGCTGTGCTACT---GGCACATCGCTCATGCTGCC 2105  
 QY 619 AlaValProGlyLysGluGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
 DB 2106 GTGAACCCCTGGTGGCGCGGACCTGCTGGTGGCTGTTCCCGAGGCGGCTCTCTCCAC 2165  
 QY 638 SerThrAsnGlyGlySerSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
 DB 2166 TCGACCGACTTGGCGCTCTGTTACAGGGGTAGGTACCGCCCAACCGGACCTCTGTTAGC 2225



QY 657 ValGlyPheGlyIysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
Db 2226 GTGGGCGCCCAAGTCCAAAGTCGACGCGCAAGAGGCTAGCGCGCTCCGCGGTCTTC 2285  
QY 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspCysGly 690  
Db 2286 ATCTGGGGCAACGACCAAGCTGGAGCGACATCGGCTGTACCGCTCCGACGCAACGGC 2345  
QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
Db 2346 AGCACCTGGACGCGCGTCAATGACCAAGAGGACACAACTACTCGGCG-----CCACCATGATC 2402  
QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
Db 2403 GAGCGCGACCCCAAGGTCTAGCGCGCGGTGTATCTAGCGCAAGACCGCGCGGTATCGTG 2462  
QY 731 TyrGlyAspIle 734  
Db 2463 TAGCGCGACCTT 2474  
RESULT 12  
US-10-395-241-17  
; Sequence 17, Application US/10395241  
; Publication No. US20040038367A1  
; GENERAL INFORMATION:  
; APPLICANT: YAOI, Katsuro  
; APPLICANT: MITSUISHI, Yasushi  
; TITLE OF INVENTION: NOVEL XYLOGLUCAN OLIGOSACCHARIDE-DEGRADING ENZYME, POLYNUCLEOTIDE  
; TITLE OF INVENTION: ENCODING THE SAME, AND METHOD OF PREPARING THE ENZYME  
; FILE REFERENCE: Q73756  
; CURRENT APPLICATION NUMBER: US/10/395,241  
; CURRENT FILING DATE: 2003-03-25  
; PRIOR APPLICATION NUMBER: JP 2002-83433  
; PRIOR FILING DATE: 2002-03-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 17  
; LENGTH: 2481  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Xyloglucan Oligosaccharide-Degrading Enzyme with Histidine Tag  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2481)  
US-10-395-241-17  
Alignment Scores:  
Pred. No.: 5.95e-97 Length: 2481  
Score: 1114.00 Matches: 281  
Percent Similarity: 50.2% Conservative: 123  
Best Local Similarity: 35.0% Mismatches: 280  
Query Match: 27.6% Indels: 120  
DB: 7 Gaps: 31  
US-09-917-376-3 (1-740) x US-10-395-241-17 (1-2481)  
QY 6 TyrThrTrpSerAsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsn 25  
Db 13 TAGAGTTCAAGAATGTGCGGATCGCGCGCGGGGTACATTACCGGGATTGTGCGGCAC 72  
QY 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAsp 45  
Db 73 CCAAGACCAAGACCTGTGCTAGCGCGCGGACGACATTCGCGCGCGGTACCGCTGGGAC 132  
QY 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
Db 133 GCAGGACAGTCCCAAGTGGATCCGCTCAACAGATTATTCAGGCGGAGGACATGAACATT 192  
QY 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
Db 193 ATGGGCACCGAGTCGATCGGCTGGAGCCCAACCAACCCCGACAGGCTGTACCTCGCGCAG 252

QY 86 GlyMetTyrThr---AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
Db 253 GGGCGCTATGTGCGCGACAGCTGG-----GCGCGCTTCTATGTGTCGGAAGAC 300  
QY 105 GlnGlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGlyAsnMetProGly 124  
Db 301 CGCGCCAGTCTGTACAACTTACAGAGTCGCGGTTCGCGATCGGCGCCCAACACATGGGA 360  
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144  
Db 361 CGCAACATGGAGCGGCTGCTGTCAACCGGTCACTCGAACGAGTCTGGATGGGT 420  
QY 145 AlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet 164  
Db 421 ACGCGTACA---GAGGGTATCTGGAAGATTCTGGACCGCGCCCAAGACCTGGACAACGTC 477  
QY 165 ThrAsnProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
Db 478 ACGTCCATCCCGGACGGCTTC-----ACCAACGGTATCGGATACACG 519  
QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerSerLeuGln 204  
Db 520 TCG-----GTCATTTTCGACCCC-----GAA 540  
QY 205 AlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSerArg 224  
Db 541 CGTAATGSCACCATCTACGCGAGCGGAGTGCCTCCCGCAGGGC---ATGTACGTACGCAC 597  
QY 225 AspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIle----- 241  
Db 598 GACGCGGTGTCTCGTGGGAGCCAGTGGCGGCGCAGCGCTCCAGCTGCGTCAACAGGACC 657  
QY 242 -----ProHisLysGlyVal 246  
Db 658 ACGGCGGTTCGCGGCAAGAACGCCGCTGATGCGCGCGCAGCCCATGAAGTCGCT 717  
QY 247 PheAspProValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAsp 266  
Db 718 CTCACCCCC-----AATCTCTCTAGCTGATTAGCGGACTACCTGGTCCATGGGGC 771  
QY 267 GlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArgIle--- 285  
Db 772 GTCACGTTCGCAAGTCTGCGCGCAGAACCGCACCTCGGCGCGCTGGGACGACATTACT 831  
QY 286 -----SerProValProSerThrAspThrAspThrAlaAsnAspTyrPheGly 299  
Db 832 CCGCGCGTCGCAACTCGTCCCTCCCGCTACAAACCAACAGACGTTCCCTCGCGCGGA 891  
QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
Db 892 TTTTGGCGTCTCAGCGTCAGCGCACCAACCCCAACCGTCTCGTCTATCACC---CTC 948  
QY 320 SerTrpTrpProAspThrIle-----IlePheArgSerThrAspGlyAlaThr 336  
Db 949 GACGCGACCCCGGACCGCCCTCGACAGCATCTACCTCTCAACCGATCGCGCGGACC 1008  
QY 337 TrpThrArgIleTrpAspTrpThrSer-----TyrPro 347  
Db 1009 TGGAAAGACGTCACCCAGCTCTCGTCCCGCTCAACCTCGAAGGTAACTGGGGCCACCG 1068  
QY 348 AsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVal 367  
Db 1069 ACTAACCGCGCGCGGTAC---AAGACGCGACAGCTGTGTGGTGGCTGACTTCAACAAAC 1125  
QY 368 GlnPro-----AsnProProValProSerPro-----LysLeuGly 379  
Db 1126 GGTCCCAAGTGGGGGATACGCTGCGCGCACGTCAGCTCCCGCGCTCACCAGTTTGGC 1185  
QY 380 TrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeuTyrGlyThr 399  
Db 1186 TGGTGGATGAGCGGTGTGCTTATCGATCCGTTCAACCCCGACGACCTGTATCGGCAG 1245

QY 400 GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLys-----TyrAspSerGly 415  
DB 1246 GGGGGACCATCTGGCGACCGACACAGCTCTCCCGTTCGAGAGGACTGG----- 1296  
QY 416 GlyGlnIleHisIleAlaPro-----MetValLysGlyLeuGluGluThrAla 431  
DB 1297 -----GCGCCGAGCTGGTACCTCCAGATCGAGCGTATCGAGAGNATCGG 1341  
QY 432 ValAsnAspLeuIleSerProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeu 451  
DB 1342 ATCTGTCTCGCTCGCTCGCCCAAGAGCGCGCGCTCTCTGTCTCGGCATCGGTGACATT 1401  
QY 452 GlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal 471  
DB 1402 AGCGCATGAAGACGACGACCTCACCAAG-----CCCCAGAGATGTTTGGTGGCCCGAC 1458  
QY 472 PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAla 491  
DB 1459 TTTCTCAACCTCGACGATCGAGCTCGGGGCACTTCCCAAGTTGTCTCGGCCCC 1518  
QY 492 GlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGly 511  
DB 1519 GGATCTCTGGGACAGTAGTACGACGCGCGTGGCGCGTGTACGCGACTGACGCG 1578  
QY 512 GlyLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrThr----- 526  
DB 1579 GGAGACGCTGGGACCATCTTCCCTACCTCGCTCGCTCGCATGAACGCGGACCTATCCAG 1638  
QY 527 GlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaPro-----Gly 544  
DB 1639 GCGACGAGATTGCGATCGCGCGCGGCGAGCAGATCGTGTGTGTCGACCAAGCTTGAC 1698  
QY 545 AspProGlyGlnProValValThrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGln 564  
DB 1699 GAGCAGCGCTGGGACCGTGTACTCGCACACTATGGCAAGACGTGTCT----- 1749  
QY 565 GlyValProAla-----AsnAlaGlnIleArgSerAspArgValAsn 578  
DB 1750 ---GTTCCGGTGGCGACCTGAAGGCCCGCAGACTGCCAATGTCTCTCGACAAAGTCCAG 1806  
QY 579 ProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVal 598  
DB 1807 GATGSCAGCTTCTACGCTACCGATGGCGGCAAGTTCTTCTGCTCGACCGCGGGGGAAG 1866  
QY 599 ThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHis 618  
DB 1867 TGTATGCGCGCAAGGCGCGGACTTGTCACT-----GGCAGATCGCTCATGCTCTGCC 1920  
QY 619 AlaValProGlyLysGlyAspLeuTrpLeu---AlaAlaSerSerGlyLeuTyrHis 637  
DB 1921 GTGAACCTCGGTGGTGGCGGCGAGCTGTGGTGTCTTCCGAGGCGGTCTCTTCCAC 1980  
QY 638 SerThrAsnGlyLysSerTrpSerAlaIle---ThrGlyValSerSerAlaValAsn 656  
DB 1981 TCGACCGACTTGGCGCTCTGTTCCAGAGGTAGGTACCGCAACGCGACCTCGTGAGC 2040  
QY 657 ValGlyPheGlyLysSer-----AlaProGlySerSerTyrProAlaValPhe 672  
DB 2041 GTCGCGCGCCCAAGTCCAAGTCGAGCGGCAAGAGGCTAGCGGCCCTCGCGGTCTTC 2100  
QY 673 ValValGlyThr-----IleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGly 690  
DB 2101 ATCTGGGGCACCGAAGCTGGAGGACATCGGCTGTACCGCTCGAGCAACAGCGC 2160  
QY 691 ThrThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIle 710  
DB 2161 AGCATTGGAGCGCGCTCAATGATGACGAGGACACAACTACTCGGGC---CCCAACATGATC 2217  
QY 711 ThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleVal 730  
DB 2218 GAGGCCAGCCCAAGGCTTAGCGGCGGTGTATCTAGGACAGAACGCGCGGTATCGTG 2277  
QY 731 TyrGlyAspIle 734

DB 2278 TACGCGGACCTT 2289  
||| |||:|

## RESULT 13

US-10-420-191-1  
; Sequence 1, Application US/10420191  
; Publication No. US20040067569A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20040067569A1ozymes Biotech, Inc.  
; APPLICANT: Rey, Michael W.  
; APPLICANT: Zaretsky, Elizabeth J.  
; APPLICANT: Haas, Jeffrey A.  
; TITLE OF INVENTION: Polypeptides Having Xyloglucanase Activity And Nucleic Acids  
; TITLE OF INVENTION: Encoding Same  
; FILE REFERENCE: 10210-200-US  
; CURRENT APPLICATION NUMBER: US/10/420,191  
; CURRENT FILING DATE: 2003-04-18  
; PRIOR APPLICATION NUMBER: US 60/373,987  
; PRIOR FILING DATE: 2002-04-19  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 5698  
; TYPE: DNA  
; ORGANISM: Tricoderma reesei  
US-10-420-191-1

## Alignment Scores:

Pred. No.: 9,498-74 Length: 5698  
Score: 878.50 Matches: 334  
Percent Similarity: 29.9% Conservative: 119  
Best Local Similarity: 22.1% Mismatches: 270  
Query Match: 21.8% Indels: 793  
DB: 7 Gaps: 52

US-09-917-376-3 (1-740) x US-10-420-191-1 (1-5698)

QY 7 ThrTrpSerAsnValAlaIle---GlyGlyGlyGlyPheValAspGlyIleValPheAsn 25  
DB 242 TCATGGAGACAGCTCAAGCTCGCGCGCGCGCGCTTCTGTCGCCGATCATCTTCCAA 301  
QY 26 -----GluGlyAlaPro----- 29  
DB 302 AGTACCTTCTTTCAGTTCGAGCGCGCGCGCGCGAGCAGCGGCGGTAGTAGAAGTT 361  
QY 30 -----GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp--- 44  
DB 362 CCCAAGACAAAAGCGGTAGCATATGCAGAACAGATATTGGCGGGCTGTACCGCCTCAA 421  
QY 45 -----AspAla 46  
DB 422 GGGTCTCTTTTCCGATCGTATAGCTGTCTGTATTAACCGCCGACATGGCGGAGTTC 481  
QY 47 AlaAsnGlyArgTrpIleProLeuLeuAspTrpVal-----GlyTrp----- 60  
DB 482 GCGGACGACTCATGGACCGCGCTCAGGATGGATGCTGTATATCCCGCTGGCACAAG 541  
QY 61 -----AsnAsnTrpGly-----Tyr 65  
DB 542 CGGCTGTCTGAGTACCTGCGCGCGAGTGCCTTACCTAACGACTATTACGCGCGACCGTTC 601  
QY 56 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnIleValTrpAlaIleVal 85  
DB 602 TGGGCGATCGAGCTGTGCGCTTGTATCGCAGGACGATCAAAAGGTGTATCGCGCAGTG 661  
QY 85 ----- 85  
DB 662 ACCCGTAGTTCGACAAAGCGAACTAGCGCTCTGCTAGTGTTCACATACGCGCTCAC 721  
QY 86 GlyMetTyrThrAsnSerTrpAsp----- 93  
DB 722 GGCATGTATACAAACAGCTGGTC-TGTGATGTCTCTCAGATCTAGACCTATGATTGGACG 780

QY 93 ----- 93  
Db 781 GCGTACATATGCTTGTGACGACAGACACTACAGAGAGTCTAGATCTCGATAAAGCTGAC 840  
QY 94 -----ProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
Db 841 CTGACATTGGCCATATAGGGATCCGAGTAATGGAGCCATCATTCCTGCTAGACCCG 900  
QY 105 ----- 105  
Db 901 GACTGTAAACGGTATATATCCCTAGGCTCATTTACCTCGGTAGTAAGCGAGCTGGCG 960  
QY 106 GlyValThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
Db 961 GCGCAACGTTGGTCTTCAACAACTTGCCCTTCAAGTCGGGGGTAAATGCCAGGAGCG 1020  
QY 125 ----- 125  
Db 1021 CCGCGTTGACACAGGAAGTGGTTGAACGGGAAGTTTTCAGCCCCCATTTAGGTCTCTGGC 1080  
QY 126 GlyMetGlyGluArgLeuAlaValAspProAsnAspAsnIleLeuTyPheGlyAla 145  
Db 1081 GGAGCCGGAGAGCGTCTGGCTGTGATCGATCCGCGCAACTCCAAACATCATCTACTTTGGTCT 1140  
QY 146 Pro----- 146  
Db 1141 CTTCCGCGCTCTCGCAGACCGACAGCTAGGCGGTTGAGGTTGTAGTAGATGAACACCA 1200  
QY 147 ---SerGlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMet--- 164  
Db 1201 CGTTCAGGAACGCGCTCTGGAAGTCTACGGAGCGCGCGTGACCTTTTCCAAAGGCTCG 1260  
QY 165 -----Thr 165  
Db 1261 GCGAGTCCTTTTCGCGGAGACCTTCAGATGCTGCGCGCGCACTGGAAAGGTTCCAGAGC 1320  
QY 166 AsnPheProAspValGlyThrTrpIleAlaAsnProThrAspThrThrGlyTyGln--- 184  
Db 1321 TCGTTACGGCAACTGGGAGCTATACATCCAGACCGAGTGATTCACACGCGCTACACAGC 1380  
QY 185 -----Ser 185  
Db 1381 AGCAAGTGGCTTGACCTGCATGTAGGTCTGGGCTCATAAGTTGCCGATGTTGTGC 1440  
QY 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAla 205  
Db 1441 GACAAGCAAGGACTCATGGGTACGTTCGATCTCAACAGCAGCAGCCGGGGAGCC 1500  
QY 205 ----- 205  
Db 1501 CTGTTCCTTCTGAGTACACCCCAATCAAGCTGAGTTGGTGGTCTGCTGGCCCTCGG 1560  
QY 206 SerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrp--- 222  
Db 1561 ACGTCTCGTATCTTGTGGACGCGGTGATTAACATCATCTCTTCTAGTCTATGTGAGCAGC 1620  
QY 223 -----SerArg 224  
Db 1621 TGCAGAGCATAGAAACACACGTCGCGACTATTGTAGTGACGAAGTACAGATCTCGTGC 1680  
QY 225 AspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGlyPheIleProHis--- 243  
Db 1681 AATGCCGGCTCCACGTGGAGTGTGTACCGGGGACGCCAGGAAATACTTTCTCTACAAG 1740  
QY 244 -----LysGlyValPhe 247  
Db 1741 TTACGCGCGAGGTGCACCTCAGCATGCGCCCGTCCGTCCCTTTATGAAGAGGTGTTCC 1800  
QY 248 -----AspProValAsnHisValLeuTyIleAlaThrSerAsnThrGlyGlyPro 264  
Db 1801 GCGAAACTGACGACGAGAGAGGCGCTTGTATCTGACCTATTTCGATGGCAGAGGCGCG 1860  
QY 264 ----- 264

Db 1861 CGTTTTCAGCTCGGTCTCTTCCGGAACATAGACTGGATAAGGCTACCGTGTCCCGGC 1920  
QY 265 TyrAspGlySerSerGlyAspValTrpLysPheSerValThrSerGlyThrTrpThrArg 284  
Db 1921 TATGATGGCACACTTGGCTCAGTGTGGAGGTACGACATTCAGGGGGAACCTTTGGAAAGAC 1980  
QY 285 IleSerProValProSerThrAspThr----- 293  
Db 1981 ATACTACCGTGAACCGAGTCACACCTCCATCTGTAAACGTCGCCCTTGAACCTTTCTG 2040  
QY 294 -----AlaAsnAspTyPheGlyTyPheGlyLeuThrIleAsp--- 306  
Db 2041 ATCACCCTGTCTCTGATCAGATCTATCTTTGGCTTTGGCGGCTTGGCCTCGATTTG 2100  
QY 307 -----Arg 307  
Db 2101 TAGTGGGGCAGACACCTAGTCTAGATATGAACCGAAACCGCGGAAACCGGAGCTAAAC 2160  
QY 308 GlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThr----- 325  
Db 2161 CAAAGCCAGGAACCTTGTGTGTTGCTTCTTTGAACCTCTTGTGGCCAGATGCTCAGCNG 2220  
QY 326 -----IleIle 327  
Db 2221 GTTTTCGGTCTTGGGAACCAACAGAAACTTTGAGAACCCGCGTCTACGAGTCCGAC 2280  
QY 328 PheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTrpPro 347  
Db 2281 TTTCCGTCGACCGACTCTGGGACCAACATGGAGCCGATCTGGCGTGGCGAGCTATCCG 2340  
QY 348 AsnArgSer---LeuArgTyPheValLeuAspIleSerAlaGluPro----- 361  
Db 2341 AAAGCCAGCTGGCTGAGACC-CTGTGTGTACCTCGGGCTAGACCCGCGCTCGATAGG 2399  
QY 362 -----Trp 362  
Db 2400 CACTGAGACCTATTACTACAGCATCTCAGTGAGTCACTCTTAAACGATCCGATGCGAATGG 2459  
QY 363 LeuThrPhe-----GlyValGlnProAsn----- 370  
Db 2460 CTGACTCTGGATATGATGCTGATAGTCACTCAGTGAGAAATGCTAGGCTACGTTACC 2519  
QY 371 -----ProProValPro----- 374  
Db 2520 GACTGACTCTGCTTCAAGACTCCCAAGACACCGTGGATCAAGAACAACTTTATCGATGTG 2579  
QY 374 ----- 374  
Db 2580 ATGACTGAGACGAAGTTCTGAGGCTTTCGTCGACCTAGTCTTCTTGTGAAATAGCTACAC 2639  
QY 375 -----SerPro-----LysLeuGlyTrpMetAspGluAlaMet 385  
Db 2640 TCGAGCGAGTCAACCGTCCGATGGTCTCATCAAGCGCCTCGGCTGGATGATTTAGTCTCTC 2699  
QY 386 Ala----- 386  
Db 2700 GGCTCGCTCAGTGGCAGGCTACAGAGTAGTTCCGGGAGCCGACCTACTAAGTACAGAGAG 2759  
QY 387 ---IleAspProPheAsnSerAspArgMetLeuTyGlyThrGlyValaThrLeuTyAla 405  
Db 2760 CAGATTGACCCACCGACAGCAACCACTGGTCTCTACGGCACCGGATACACAATCTTTGGC 2819  
QY 406 Thr----- 406  
Db 2820 GTCTAACTGGTGGTGTCTGTTGGTGAACCGAGATGCGTGGCCTTACTGTTAGAAACCG 2879  
QY 407 ---AsnAspLeuThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLys 425  
Db 2880 CGCCAGCATCTCACCACACTGGGACACCGCCCAACATGTGTCAATCCAATCACTGCGCAGC 2939  
QY 426 Gly----- 426

Db 2940 GCGGTGTAGAGTGGTTCACCTGTGCGCGGTGTGTACACAGTTAGTTAGTACCGCTGTG 2999  
Qy 427 ---LeuGluGluThrAlaValAlaAsnAspLeuIleSerProSerGlyAlaProLeuIle 445  
Db 3000 CGCATCGAGGAATTCTCGTCCAGAACCTGCGCTCTGCACCGCGGGAAGCGAGCTATTG 3059  
Qy 446 Ser----- 446  
Db 3060 GCGTAGCTCTTAAAGACGAGCTCTGTGACCGGAGAGAGTGGCGCGCTTCCTCGATAAC 3119  
Qy 447 ---AlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaVal- 462  
Db 3120 CCGCAGTCGGAGAGCAACAGGCTTACCTTGGCCAGCAAGAACCTCGGACATCG 3179  
Qy 462 ----- 462  
Db 3180 GCGGTGACGCTCTGCTGTTGCCGAAGTGGAACCGTGTCTTTGCTGGAGCCCTGTAGC 3239  
Qy 463 ProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAla- 481  
Db 3240 GCGCAGAGCGTCTGGCAACGCCACATGGGCCACCTCGACGAGCGTCACTACGCCCGG 3299  
Qy 482 -----Glu 482  
Db 3300 AGCGTCTGCCAGACCGGTGCGGGTGTACCGGTGGAGCTGCTCGACAGTATGCGGCC 3359  
Qy 483 LeuAsnProSerIleValAlaGlySer----- 493  
Db 3360 TACTCGGTCAAGAGCGTCTCGCGTGGCAACACCGCGCGGCAACACAGGTGGCCATC 3419  
Qy 494 ---PheAspProSerSerGlnProAsnAspArgHisAlaAlaPhe- 507  
Db 3420 TTGAGCCAGTTCTCGCAGCAGCGCAGCGGTGTGGCGCGGTGTGTCCACCGGTAG 3479  
Qy 508 SerThrAspGlyGlyAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGly 527  
Db 3480 ACGTCCGACGCGCGCGACGTGGAGCATCGACTACGCGCGCGACACGTCCATGAACGCG 3539  
Qy 528 Gly----- 528  
Db 3540 GCGAGCTGCGCGCGCGTGCACCTGTAGTGTATGCGCGCGGTGTGAGTACTTGGCG 3599  
Qy 529 ---ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGly 547  
Db 3600 GCGACGTTGGCTTATCGCGCGACGCGACAGATCTCTGTGTGACCGCTCTGTCGCGC 3659  
Qy 548 -----GlnPro 549  
Db 3660 GCGTGCCACCGGATAAGCGCGTGCCTGTGTAGGAGACAGCTGGCGGAGCGCG 3719  
Qy 550 ValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsn 569  
Db 3720 CTGCAGCGCTCG-CACTTCCAGGCGAGCTTTCCTCGCTCGAGCGCTCGCGCGCGC 3776  
Qy 570 AlaGln----- 571  
Db 3777 GCGAGCTGCGAGGTCAGGTCCCGTTCGAAACGAGGAGAGCTCGGACGGGCGCGC 3836  
Qy 572 ---IleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThr 589  
Db 3837 CGGCTCATCGCTCGGACAAAGAACCAACAGCGTCTTCTACGCGCGGTTCGGATCGACC 3896  
Qy 590 Phe----- 590  
Db 3897 TTTTAGTAGCGAGCTGTTCTTCTGTTGTTCGAGAAAGATGCGGCGGAGCTAGCTGG 3956  
Qy 591 ---TyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSer 609  
Db 3957 AAAACGTCTCAGAGACACCGGACGAGCTTC-----ACGCGCGG-CCCAAG 4004  
Qy 610 SerGlyAlaValGlyValMet----- 616  
Db 4005 CTGGCAGCGCAGGAGTGCAGTCTGTTGGCGGTGTCAGTGTGCGCGCGCGCGGTTTC 4064

Qy 617 -----PheHisAlaValProGlyLysGluGlyAspLeu 627  
Db 4065 GACCGTCCGFCCTCGATCCGGGATATCGTGTCTCACCCGACCCGCGGCGACGTTG 4124  
Qy 628 TrpLeuAlaAlaSerSer----- 633  
Db 4125 TATGTCTCGACCGCGCTAGGCCCTATAGCAGAGTGGGTGTGGCGCCCGTGCAC 4184  
Qy 634 -----GlyLeuTyrHisSerThrAsnGlyGlySer- 644  
Db 4185 ATACAGAGTGGCTCTCGGCATATTCCTCCACAGACTCGGCGACGACCTTTGGCCAA 4244  
Qy 644 ----- 644  
Db 4245 GTCTCCACCGCCTGAAAGCGGTATAAGCGAGGTGTCTAGCGCGCTGTGAAACCGGTT 4304  
Qy 645 ---TrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662  
Db 4305 CAGAGTGGCGGAGTCCAAACACTACAGATGCGCCTGGGTGTGGGCTCAGGCTCGAAC 4364  
Qy 663 -----AlaProGlySerSerTyr 668  
Db 4365 TCGAACCTGTATGCTCTGTTGTGGATGTCTAGCGGAGCCACACCGAGTCCGAGCTTG 4424  
Qy 669 ProAlaValPheValValGlyThrIleGlyValThrGlyAla- 686  
Db 4425 ACCTTGGACATACGATCGGCACC-----GGCCGTCAGGGGCTCGCCTCTACGCCAGT 4478  
Qy 687 AspAspCysGlyThrTrpVal----- 694  
Db 4479 GGAGACAGCGCGCTCTCTGGAAGCGTGGCGGCGAGTCCCGAGCGAGATGCGGTCA 4538  
Qy 695 -----LeuIleAsnAspAspGlnHisGlnTyrGly- 704  
Db 4539 CCTCTGTCCCGCGGAGGACCTCGACATCCAGGGCTCCCGAGGGCTTCGGCTCCATCGAC 4598  
Qy 704 ----- 704  
Db 4599 AGCACCAGGTCCCGCGGAGCGCTGTAGTCCCGAGGGTCCCGAGCGAGTAGCTG 4658  
Qy 705 AsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThr 724  
Db 4659 TCGTGTTCAGCGCGCGTCCGCGAGCAGCAGCGCGG-----CAAGTCTACGTGGGCACC 4712  
Qy 725 AsnGlyArgGlyIleValTyrGlyAspIleGlyAlaPro 738  
Db 4713 AACGCGCGGGCGTCTTTTACGCTCAGGCGTCTGTGGCGGCC 4754

## RESULT 14

US-10-653-047-7511  
; Sequence 7511, Application US/10653047  
; Publication No. US20040229367A1  
; GENERAL INFORMATION:  
; APPLICANT: Randy M. Berka  
; APPLICANT: Michael W. Rey  
; APPLICANT: Jeffrey R. Shuster  
; APPLICANT: Sakari Kauppinen  
; APPLICANT: Ib Groth Clausen  
; APPLICANT: Peter Bjarke Olsen  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; TITLE OF INVENTION: Expression  
; FILE REFERENCE: 5849.200-US  
; CURRENT APPLICATION NUMBER: US/10/653,047  
; CURRENT FILING DATE: 2003-08-29  
; PRIOR APPLICATION NUMBER: US/09/533,559  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 09/273,623  
; PRIOR FILING DATE: 1999-03-22  
; NUMBER OF SEQ ID NOS: 7860  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7511  
; LENGTH: 1103

```

; TYPE: DNA
; ORGANISM: Tricoderma reesei
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1103)
; OTHER INFORMATION: n = A,T,C or G
US-10-653-047-7511

Alignment Scores:
Pred. No.: 1,01e-27 Length: 1103
Score: 392.50 Matches: 114
Percent Similarity: 45.0% Conservative: 36
Best Local Similarity: 34.2% Mismatches: 125
Query Match: 9.7% Indels: 58
DB: 8 Gaps: 10

US-09-917-376-3 (1-740) x US-10-653-047-7511 (1-1103)

Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321
Db 3 GGCCCTGGCCCTCGATTGCAAAAGCCAGGAACCTTGTGTGCTTCTTGAACCTCTGG 62
Qy 322 TrpProAspThrIleIlePheAtgSerThrAspGlyGlyAlaThrTrpThrArgIleTrp 341
Db 63 TGGCCAGATGCTCAGCTGTGTTCGTCACCACTCTGGACATCGAGCCGATCTGG 122
Qy 342 AspTrpThrSerTyProAsnArgSerLeuArgTyValLeuAspIleSerAlaGluPro 361
Db 123 CGGTGGCGGAGTATCCGATGAGCTATTTACTACAGCATCTCAACTCCCAAGCACCG 182
Qy 362 TrpLeuThrPheGly---ValGlnProAsnProProValProSer-----Pro 376
Db 183 TGGATCAAGAACAACTTTATTCGATGTGACGAGCGAGTCACCGTCCGATGCTTNATCAAG 242
Qy 377 LysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAspArgMetLeu 396
Db 243 CGCTTCGGCTGGATGATGTAGTCTNTCGAGATGACCCACCGCAACNACTGGCTT 302
Qy 397 -TyGlyThrGlyAlaThrLeuTyAlaThrAsnAspLeuThrLysTrpAspSer----- 414
Db 303 TTACGGACCGGATGACATNTTGGCGGCCAGCATTTCCCACTGGACAGCGGCC 362
Qy 415 -----GlyGlyGlnIleHisIleAlaProMetValLys-GlyLeuGluThrAlaV 432
Db 363 ACAATGTGGTCAATCCATTACTTGGCAGAGCGGATTGGAAGGAATT-----TTCCGT 416
Qy 432 alaAsnAspLeuIleSerProProSerGly-----Pro----- 441
Db 417 TCAAGGACCTGGCCCTTTTCACCCGGGGGGAAGCGAGCTTTTGGCCGCAAGTCCGAGACG 476
Qy 442 -----AlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspV 459
Db 477 ANCAACGGGTTTACCTTTTCCCGCAGCAGAAACGACCTTGGG----- 517
Qy 459 alThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThr-SerVal 478
Db 518 -----ACATTGGCCGACAGCGTTTGGCAACTCCACATCGGCCACCTCGACGAAGCTC 572
Qy 479 AspTyrala--GluLeuAsnProSerIleleValArgAlaGlySer----- 493
Db 573 GACTACCGCGGAACCTCGTCAAGAGCGGTTTCGTCGCGTCCGCAACACCGCGCGGACG 632
Qy 494 -----PheAspProSerSerGlnProAsnAspArgHisValAlaPheS 508
Db 633 CAACAAGTGGCCATTTTTCGAAACGGCGGGCGGCGAGCTNGGAA-----GCAATTC 686
Qy 508 erThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyG 528
Db 687 GAACATACGCT-----GNTCCGACACCGCTTCCATTGAACCGCG 728
Qy 528 lyThrValAlaIleSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyG 548
Db 729 GCGCGGTGGCTATTTCGGCCGACGCGGACACGATCTCTGTCGACCGCTCGTCCGCGC- 787

```

---

```

Qy 548 lnProValValTyAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProA 568
Db 788 -----GTGCAGCGCTCGCAGTTCCAGGGCAGCTTTGCTCCTCTCGAGCTGCCG 839
Qy 568 laAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyralaLeuSerAsnG 588
Db 840 CGGGCGCGCTCATCGNCTCGGACAAAGACCAACAGCGTNTTCTACGCCGCTCCGGAT 899
Qy 588 lyThrPheTyArgSerThrAspGlyGly 597
Db 900 CGACCTTTTACGTCAGCAAGGACACCGCGC 928

RESULT 15
US-10-246-330-3
; Sequence 3, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; FILE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7407
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7404)
US-10-246-330-3

Alignment Scores:
Pred. No.: 9,48e-12 Length: 7407
Score: 238.00 Matches: 205
Percent Similarity: 32.6% Conservative: 88
Best Local Similarity: 22.8% Mismatches: 314
Query Match: 5.9% Indels: 292
DB: 6 Gaps: 50

US-09-917-376-3 (1-740) x US-10-246-330-3 (1-7407)

Qy 11 ValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAlaProGly 30
Db 3250 GTCAACCTGAGCAATGGCAGCAGCTCAGCGC-----ACGCCGAGCGCGC 3297
Qy 31 IleLeuTyValArgThrAspIleGlyMetTyArgTrpAspAlaAlaAsnGlyArg 50
Db 3298 AGCACGGTGATCTTCACCGAC-----GGCAACGCGCAAT 3330
Qy 51 TrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyT----- 65
Db 3331 CCGATCGCCGAGGTTCACCGCGCAGCGGCAACTCGACCTACACCCCGTCCACGCGC 3390
Qy 66 -----AsnGly---ValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp 82
Db 3391 ATCCCAACCGTACTGTGTCAAGTGTGTGTCGCCAGGACGCTCCGTTAACGAGCGCG 3450
Qy 83 AlaAlaValGlyMetTyThrAsnSer-----TrpAspProAsnAsp 96
Db 3451 CCGCGCAGCGTGACCGTCGATTCAGCGCGCGCGCGCGGTGATCAACCCGAGCAAC 3510
Qy 97 GlyAlaIleLeuArg---SerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuPro 115
Db 3511 GCGCTCGTTCATCAGCGGCAACCGCGAGGCGGTGCGACGCTCACC----- 3561
Qy 116 PheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlyLeuAlaValAspPro 135

```

Db 3562 ---GATCGCGGCAACCCGATAGGCGAGGTACCGCGCAC--- 3600  
QY 136 AsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThr 155  
Db 3601 ---GGCAGCGGCAACTGGAGCTTCACG 3624  
QY 156 AspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAla 175  
Db 3625 CGG---GGCAGCGCGGTGGCC 3642  
QY 176 Asn---ProThrAspThrThrGlyTyrGlnSer 185  
Db 3643 AACGCAGCGTATCGTCGCCACGGCCACCGACCGGCAATACCGCCCGCGAGGCC 3702  
QY 186 ---AspIleGlnGlyValValTrpValAlaPheAspLysSerSerSer 201  
Db 3703 GCCACACGGTGGACGGGTGGCGCGCGCGCGGTGATCGATCCGAGCAACGGCAGC 3762  
QY 202 Leu---GlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
Db 3763 ACCATCAGCGCACCGCGGAGCGCGGCGCAAGGTGATCTCACCGACGGCAACGGCAAC 3822  
QY 219 ProVal---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
Db 3823 CGGATCGCGGAAACACACCGCGACGGCAGCGGCAACTGGAGCTTCACCGCCCGGCGCGCG 3882  
QY 238 ---ThrGly--- 239  
Db 3883 CTGGCCACGGCACGGTGGTCAACGCGCGTGGCCAGGACCGTGGCGGCAATACCGGCCCG 3942  
QY 240 ---PheIleProHisLysGlyValPheAspProValAsn 251  
Db 3943 CAGGGCAGCACTACCGTGGACGGGTGGCGCGCAACCGCTGTGGTCAATCCGAGCAAC 4002  
QY 252 HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyr---AspGlySer 268  
Db 4003 GCGAACCTGTCTCAACGGTACCGCGGAGCGCGGCGACCGTGCACCTTCACCGCACGGCAAC 4062  
QY 269 SerGlyAspValTrpLysPheSerValThr---SerGlyThrTrp---ThrArgile 285  
Db 4063 GCGAACCGGATCGGCCAGCACCGCGCGATGCGACGGCAACTGGAGTTTCACGCGCCGCG 4122  
QY 286 SerProValProSer---ThrAspThrAlaAsnAspTyrPheGlyTyrSer 301  
Db 4123 TCGCACTGCCCAACGGCACCGTGTCAACGTGACCGCGGAGCGCGCGCGGCAATACC 4182  
QY 302 GlyLeu---ThrIleAspArgGlnHisProAsnThrIleMetVal--- 315  
Db 4183 AGCCTTCGCGCTACACGACGGTGGATTCCTCGTGGCTGATCCCGCGAGTGGATCCG 4242  
QY 316 ---AlaThrGlnIleSerTrpTrpProAsp---ThrIleIlePheArgSer 330  
Db 4243 AGCAACGGTTCGGTGCATCAGCGGCACCGCGGACCGCGGCAACCCATCATC--- 4296  
QY 331 ThrAspGlyGly---AlaThrTrpThrArgIleTrpAsp 342  
Db 4297 ACCGATGGCAACGGCAACCCGATGGCCAGGTTCACCGCGGACGGCGGCAACTGGTCC 4356  
QY 343 TrpThr---SerTyrProAsnArgSerLeuArgTyrVal 354  
Db 4357 TTCCTCCAGGCATCCCGCTCGCGATGGCGACCGGTGGTCAACGTGGTGGCGCGGACGCCA 4416  
QY 355 LeuAspIleSerAlaGluProTrpLeu---ThrPheGlyValGlnProAsnProPro 372  
Db 4417 AGCAATGTCACAGTCGCGCGCGGTGATCACTGTGGATGGTGGTGGCGCGCGCGCG 4476  
QY 373 ValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSer 392  
Db 4477 GTG---ATCGATCCGAGCAACCGC 4497  
QY 393 AspArgMetLeuTyrGlyThr---GlyAlaThrLeuTyrAlaThrAsnAspLeu 409  
Db --- 411

Db 4498 ACCGAG---ATAAGCGGTACCGCGGAGCGCGCGACGGTGATCTCTCACC--- 4545  
QY 410 ThrLysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluGlu 429  
Db 4546 ---GATGGCGCGGCAAC---CCGATCGCGAGGCACCGCGCAC 4584  
QY 430 ThrAlaValAsnAspLeuIleSerPro---ProSerGlyAlaProLeuIle 445  
Db 4585 GGCAGCGGCAACTGGAGCTTCACCGCGGACCGCGTCCGCAACCGCGATGATCAAC 4644  
QY 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHis---AlaAspVal 459  
Db 4645 GCGGTGGCCACGACCGCGCGGCAATACAGCGGTTCGCGCAGCGTCACCGTCGATGCC 4704  
QY 460 ThrAlaValProSerThrIlePhe---ThrSerProValPheThrThrGlyThrSer 477  
Db 4705 ATCGCGCGCGCGCGGTGATCAATCCGAGCAATGGAGTCTGTCATCAGCGGTACG--- 4761  
QY 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySerPheAspProSer 497  
Db 4762 ---GCGGAAGCGCGGCGCGGTGATCTCACCGACGGCAACCGCGCAACCGCGATC 4812  
QY 498 SerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPhe--- 516  
Db 4813 GGCCAG---GTCACCGCGCAGCGCGACGGCAACTGGAGCTTC 4851  
QY 516 --- 516  
Db 4852 AGCGCGCGCACCGCGTGGCCAAACGGTCTGGTTCATCAATGCGGTGGCCAGGACGGCGCC 4911  
QY 517 ---GlnGlySerGluProGlyGlyValThr--- 525  
Db 4912 GGCAACAAACAGCAGTCCCAACGAGCGCACCGTCTGCTGGCGCCAGCAGCCCGGTG 4971  
QY 526 ---ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrp 541  
Db 4972 ATCGATCCGAGCAACCGGTAGCTGATCGCGGTACCGCGGAGCTGGTGCACGCGATC 5031  
QY 542 AlaProGlyAspProGlyGlnProVal---ValTyrAlaValGlyPheGlyAsnSer 559  
Db 5032 CTCACCGAGCAACCGGCAACCGGATCGCGCAGGTACCGCGCGATGGCAGCGCAAC--- 5088  
QY 560 TrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPro 579  
Db 5089 TGGAGCTTCACCGCGCGCACGCG--- 5112  
QY 580 LysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArg--- 592  
Db 5113 ---CTGTCCAATGGCAGCGTGGTCAATGCGGTGGCGCCAGGACGCTGCC 5157  
QY 593 ---SerThrAspGlyGlyVal---ThrPheGlnProValAlaAlaGlyLeuPro--- 608  
Db 5158 GGCAACACGAGCGCGCGGTTCAGCACCGGTGGACCGCGGTGGCGCGCGCGCGGTG 5217  
QY 609 ---SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGlyGlyAsp 626  
Db 5218 ATCGACCGGAGCAACGGTGTGCAACTCAGCGCACCGCGCAACCGCGGTTCGGGTGATC 5277  
QY 627 LeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSer---ThrAsnGlyGlySerSerTrp 645  
Db 5278 CTCACCGATGGCAATGGCAATCCGATCGCGCAGACCGCTTGGCGCGGCAACTGG 5337  
QY 646 SerAlaIleThrGlyVal---SerSerAlaValAsnValGlyPheGlyLys 661  
Db 5338 AGCTTCACCGCGCGCGCGGTGGCGCAACCGCGCGGTGGTCAATGCGGTGGCGCGGAC 5397  
QY 662 SerAlaProGlySerSerTyrProAlaValPheValGlyThrIleGlyGlyValThr 681  
Db 5398 CCGCGCGGCAATACGAGCGCGCGCGCGCGCACCGCGGTGGTGGTGGCGCGCGCG 5457  
QY 682 GlyAlaTyrArgSerAspAspCysGlyThrThrTrpValLeuIleAsnAspAspGlnHis 701  
Db 5458 ---CCGCTGATCAATCCAGCAAC--- 5478

|    |      |  |      |
|----|------|--|------|
| Qy | 702  | GlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgArgValTyr | 721  |
| Db | 5479 | -----GGCAGCGTGATCACCGGCACC---GCCGAGGTCGGCGCCCAAGGTG          | 5520 |
| Qy | 722  | IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIleGlyGlyAlaProSerGly    | 740  |
| Db | 5521 | ATCTCACCGATGGCAACGGCAACCCGATCGCGGAGACCCGCCCGGCGGCGGT         | 5577 |

Search completed: March 2, 2006, 20:08:50  
 Job time : 11752.8 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:24:59 ; Search time 794.098 Seconds  
(without alignments)  
2043.020 Million cell updates/sec

Title: US-09-917-376-3

Perfect score: 4036

Sequence: 1 ATTQPYTWSNVAIGGGFVD.....YIGTNGRGIVGDTGGAPSG 740

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7218512 seqs, 1096188070 residues

Total number of hits satisfying chosen parameters: 14437024

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-Q=/abss/ABSSWEB/spool/US09917376/runat\_02032006\_091502\_8293/app\_query.fasta.1  
-DB=Published\_Applications\_NA\_New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -HOST=abs03p  
-USER=US09917376@CGN\_1\_1067 -runat 02032006\_091502\_8293 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published\_Applications\_NA\_New.\*

1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                    | Description        |
|------------|-------|-------------|--------|-----------------------|--------------------|
| 1          | 192   | 4.8         | 9903   | 12 US-11-052-554A-517 | Sequence 517, App  |
| 2          | 182   | 4.5         | 11151  | 12 US-11-052-554A-525 | Sequence 525, App  |
| 3          | 180   | 4.5         | 7231   | 12 US-11-136-527-2622 | Sequence 2622, App |
| 4          | 178   | 4.4         | 9474   | 12 US-11-052-554A-526 | Sequence 526, App  |

|      |       |     |        |                         |                    |
|------|-------|-----|--------|-------------------------|--------------------|
| 5    | 175.5 | 4.3 | 7572   | 12 US-11-052-554A-527   | Sequence 527, App  |
| 6    | 168.5 | 4.2 | 15876  | 12 US-11-052-554A-660   | Sequence 660, App  |
| 7    | 166.5 | 4.1 | 2943   | 12 US-11-052-554A-401   | Sequence 401, App  |
| 8    | 165   | 4.1 | 6615   | 12 US-11-052-554A-518   | Sequence 518, App  |
| c 9  | 161   | 4.0 | 37507  | 8 US-10-522-037-2       | Sequence 2, Appli  |
| c 10 | 157   | 3.9 | 3155   | 8 US-10-793-626-3881    | Sequence 3881, Ap  |
| c 11 | 155.5 | 3.9 | 2733   | 8 US-10-793-626-3107    | Sequence 3107, Ap  |
| c 12 | 155   | 3.8 | 2595   | 12 US-11-136-527-3719   | Sequence 3719, Ap  |
| c 13 | 153   | 3.8 | 3050   | 8 US-10-793-626-3661    | Sequence 3661, Ap  |
| c 14 | 152   | 3.8 | 3748   | 8 US-10-793-626-3970    | Sequence 3970, Ap  |
| c 15 | 151.5 | 3.8 | 3600   | 8 US-10-793-626-4165    | Sequence 4165, Ap  |
| c 16 | 150   | 3.7 | 3921   | 12 US-11-052-554A-523   | Sequence 523, App  |
| c 17 | 149   | 3.7 | 4101   | 8 US-10-821-234-579     | Sequence 579, App  |
| c 18 | 149   | 3.7 | 5086   | 12 US-11-186-284-30     | Sequence 30, Appl  |
| c 19 | 149   | 3.7 | 5086   | 12 US-11-091-883-109    | Sequence 109, App  |
| c 20 | 147.5 | 3.7 | 2268   | 8 US-10-517-939-329     | Sequence 329, App  |
| c 21 | 147.5 | 3.7 | 2937   | 9 US-11-202-566-8       | Sequence 8, Appli  |
| c 22 | 147   | 3.6 | 4551   | 8 US-10-220-824-7       | Sequence 7, Appli  |
| c 23 | 146   | 3.6 | 2850   | 12 US-11-052-554A-390   | Sequence 390, App  |
| c 24 | 146   | 3.6 | 104299 | 12 US-11-000-688-1364   | Sequence 1364, Ap  |
| c 25 | 145   | 3.6 | 2145   | 8 US-10-793-626-569     | Sequence 569, App  |
| c 26 | 145   | 3.6 | 2283   | 12 US-11-052-554A-452   | Sequence 452, App  |
| c 27 | 144.5 | 3.6 | 191684 | 12 US-11-121-086-2      | Sequence 2, Appli  |
| c 28 | 143.5 | 3.6 | 5453   | 12 US-11-091-883-108    | Sequence 108, App  |
| c 29 | 143.5 | 3.6 | 153376 | 12 US-11-121-086-5      | Sequence 5, Appli  |
| c 30 | 143.5 | 3.6 | 172543 | 12 US-11-121-086-6      | Sequence 6, Appli  |
| c 31 | 141.5 | 3.5 | 3024   | 8 US-10-517-939-83      | Sequence 83, Appli |
| c 32 | 141   | 3.5 | 2425   | 8 US-10-793-626-4404    | Sequence 4404, Ap  |
| c 33 | 140.5 | 3.5 | 2868   | 12 US-11-052-554A-562   | Sequence 562, App  |
| c 34 | 140.5 | 3.5 | 4386   | 12 US-11-052-554A-662   | Sequence 662, App  |
| c 35 | 140.5 | 3.5 | 88421  | 12 US-11-205-109-1      | Sequence 1, Appli  |
| c 36 | 140.5 | 3.5 | 120096 | 12 US-11-121-086-24     | Sequence 24, Appli |
| c 37 | 139.5 | 3.5 | 4767   | 12 US-11-052-554A-659   | Sequence 659, App  |
| c 38 | 139   | 3.4 | 11447  | 12 US-11-186-284-25     | Sequence 25, Appli |
| c 39 | 139   | 3.4 | 11554  | 12 US-11-169-041-34     | Sequence 34, Appli |
| c 40 | 138.5 | 3.4 | 21706  | 6 US-09-925-065A-683197 | Sequence 683197,   |
| c 41 | 138.5 | 3.4 | 5706   | 12 US-11-052-554A-519   | Sequence 519, App  |
| c 42 | 138.5 | 3.4 | 168516 | 12 US-11-121-086-3      | Sequence 3, Appli  |
| c 43 | 137.5 | 3.4 | 4311   | 12 US-11-052-554A-524   | Sequence 524, App  |
| c 44 | 137   | 3.4 | 2155   | 12 US-11-122-329-122    | Sequence 122, App  |
| c 45 | 137   | 3.4 | 4605   | 12 US-11-102-476-3      | Sequence 3, Appli  |

#### ALIGNMENTS

RESULT 1

US-11-052-554A-517  
; Sequence 517, Application US/11052554A  
; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; PRIOR FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 517

; LENGTH: 9903

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis H37Rv

US-11-052-554A-517

Alignment Scores:

Pred. No.:

Score: 0.000114

Percent Similarity: 192.00

Best Local Similarity: 35.0%

Mismatches: 23.6%

Length: 9903

Matches: 194

Conservative: 94

Mismatches: 323

|   |      |  |      |
|---|------|--|------|
| Query Match:  | 4.8% | Indels:  | 213  |
| DB:   | 12   | Gaps:  | 43   |
| US-09-917-376-3 (1-740) x US-11-052-554A-517 (1-9903) |      |  |      |
| QY  | 10   | AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGly          | 27   |
| DB  | 7399 | AACCGGAACATCGGCGCGCAAAACATCGGCAC                             | 7449 |
| QY  | 28   | -----AlaProGlyIleLeuTyrValArgThrAspIleGly                    | 41   |
| DB  | 7450 | AACACCGGTTCGGGGCTTACACCGCGCTGTCAACACATCGGTATCGGCAACACCGGCAC  | 7509 |
| QY  | 42   | TyrArgTrpAspAlaAlaAsn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrp | 60   |
| DB  | 7510 | TACAACATCGGTGTGGCGCAACACCGGTAACTACAAC                        | 7554 |
| QY  | 61   | AsnAsnTrpGlyTyrAsnGlyVal---ValSerIleAlaAlaAsp                | 74   |
| DB  | 7555 | GGCAACACCGGCAACACACATCGGCATCGGCTGTCCGGCGACACACGATCGGGTTC     | 7614 |
| QY  | 75   | ---ProIleAsnThrAsnLysValTipAlaAlaValGlyMetTyrThrAsnSerTipAsp | 93   |
| DB  | 7615 | GGCCCGGTGAACCGCGC-----ATGCCCAATGGCCCTGTTC---AACCTGGCGCAC     | 7665 |
| QY  | 94   | ProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrPro | 113  |
| DB  | 7666 | AACAACATTGGC-----  | 7677 |
| QY  | 114  | LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaVal | 133  |
| DB  | 7678 | -----ATGGCCAACCGGGCAACTTCAACACCGGCATTTGCC                    | 7713 |
| QY  | 134  | AspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGlyLys---GlyLeuTrp | 152  |
| DB  | 7714 | AACACCGGCAACACACATCGGCTTGTTCACACCGGCAACACACATCGGCTTCGG       | 7773 |
| QY  | 153  | ArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThr | 172  |
| DB  | 7774 | CTGACCGCGAGCGGTGTTCGGCTTCAGCTCCCTGAACTCCGGCGCGGCAACACCGGT    | 7833 |
| QY  | 173  | TyrIleAlaAsnProThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrp | 192  |
| DB  | 7834 | TTCTTCAACTCCGGCACCGGCAACACCGGC-----                          | 7863 |
| QY  | 193  | ValAlaPheAspLysSerSerSerLeuGlyGln---AlaSerLysThrIlePheVal    | 211  |
| DB  | 7864 | ---TTGTCAACTCCGGCACCGGCAACACCGGCTTGTTCAACTCCGGCACCGCAACGTC   | 7920 |
| QY  | 212  | GlyValAlaAlaProAsnAsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGln | 231  |
| DB  | 7921 | GGCATCGGCAACATGGGCACCGGCGGTTCGGCTCGGCCTATCCGGCGACGACGCGT     | 7980 |
| QY  | 232  | AlaValProGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsn | 251  |
| DB  | 7981 | GGCATCGGCGGCAACAACTCGGCGAGTTTC---AACATCGGCTTGTTCATCGGCGCAC   | 8037 |
| QY  | 252  | HisValLeuTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGlySerGly-As    | 271  |
| DB  | 8038 | GGCAATGTCCGCATCGGCAACTCGGCGACCGGCAACGTCGTCATCGGCAACACCGGCAC  | 8097 |
| QY  | 271  | pValTrpIlePheSerValThrSerGlyThr-----                         | 282  |
| DB  | 8098 | GGCAACACCGGCATCGGAACAGCGGCAACTACAACACCGGCTTGTTCACCGGCGCTG    | 8157 |
| QY  | 282  | p-----ThrArgIleSerProValProSerThrAspThrAlaAsnAs              | 296  |
| DB  | 8158 | GTCAACACCGGCATCGGCAACCGGCGCAACACACACCGGCTTGTTCACATCGGCAC     | 8217 |
| QY  | 296  | pTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAl | 316  |
| DB  | 8218 | TTCAACACCGGCATCGGCAACCGGCGCACTACAACACCGGCTCTCTACAACACCGGTAGC | 8277 |
| QY  | 316  | aThrGlnIleSerTrp---TrpProAspThrIleIle---PheArgSerThr         | 331  |
| DB  | 8278 | TACAACACCGGCATCGGCAACCGCGAGATACGGCACCGCGGCTTCATCACCGGCAGC    | 8337 |
| QY  | 332  | -----AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSer              | 345  |
| DB  | 8338 | ATGAACAACGGGTCTCTGCGCGCGCCACCGGAGGCGCTCTGCGCGCAACTACACC      | 8397 |
| QY  | 346  | -----TyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaG1            | 360  |
| DB  | 8398 | ATCACCATCGAGCGACCTGCGCGGTCTCTCAATGTGCATCCCGGTCAACATCCCATC    | 8457 |
| QY  | 360  | uProTrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTr | 380  |
| DB  | 8458 | ACCG---CGGACATCAACCAATGTCTCCATCCCGGCATTAAGTTCCCCAGA-ATC      | 8508 |
| QY  | 380  | pMetAspGluAlaMetAlaIleAsp---ProPheAsnSerAspArgMetLeuTyrGlyTh | 399  |
| DB  | 8509 | ---GACCCAGCGGAACGCTGCATAGGCATCTCTCAGTGGCACCGTCTTGGCCCCGGT    | 8564 |
| QY  | 399  | rGly---AlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAspSerGlyGlnI1    | 418  |
| DB  | 8565 | CGTCCGATCACTTCATGCGGCGGACGCGTGGCGCGCGCTGGACACACCATCGAAAT     | 8624 |
| QY  | 418  | eHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSerPr | 438  |
| DB  | 8625 | TGACTTCGGCCCC-----TCGCGCGCGATCACTCAACATCGGCA                 | 8666 |
| QY  | 438  | oProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAs | 458  |
| DB  | 8667 | GCCGACGCGCTCCACCGTGATCAACATGCTGGCGCGCGCGC-----               | 8709 |
| QY  | 458  | pValThrAlaValProSerThrIlePheThrSerProValPheThrGlyThrSerVa    | 478  |
| DB  | 8710 | -----GCCGCGCGGATC-----AGCAT                                  | 8726 |
| QY  | 478  | lAspTyrAlaGluLeuAsnProSerIleValArgAlaGlySerPheAspProSerSe    | 498  |
| DB  | 8727 | TCCGATCATCAGCTTCGCGCGCAGC-----CCCGGCTTCTCAACCGCACAC          | 8774 |
| QY  | 498  | rGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAsnTrpPheGlnG1 | 518  |
| DB  | 8775 | CGGCGC-----TCGTGGCGCTTCTCACTCGGGGTCTGG                       | 8810 |
| QY  | 518  | ySerLupProGlyGlyValThrThrGlyGlyThr-----ValaAla               | 532  |
| DB  | 8811 | CAGCGCATCGGCTGTCTGAACATTCGCAACAACTCGGCGCTCTCAACTCGCCACTAG    | 8870 |
| QY  | 532  | aSerAlaAspGlySerArgPheVal-----TipAlaProG1                    | 544  |
| DB  | 8871 | CAGCATGGGAATTCGGGCTTCCAAACTATGGGTGCGTGCAGTGGGTGGCG-----      | 8925 |
| QY  | 544  | yAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSer-----Tr        | 560  |
| DB  | 8926 | -----AATTGGGCAACAGCATCTCGGCGCATCTA                           | 8954 |
| QY  | 560  | pAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLy | 580  |
| DB  | 8955 | CAACACCGGCTTGGGAGCACCGGCAAAATGTC-----TCGGGCTTCTCAACATCGG     | 9005 |
| QY  | 580  | sThrPheTyrAla-----LeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyVa  | 598  |
| DB  | 9006 | CACCAACTGCTGGGTGGTTCAGAACCGGCGC-----                         | 9039 |
| QY  | 598  | lThrPheGlnProValAlaAlaGlyLeu-ProSerSerGlyAlaValGlyValMetPhe- | 617  |
| DB  | 9040 | -ACGAGACGACCTTCAGCTGGGCTGGCCCAACCTCGGTTCT---GGAACTGGGTAG     | 9095 |
| QY  | 618  | -----HisAlaValProGlyLysGluGlyAspLeuTrpLeuA                   | 630  |
| DB  | 9096 | CGCAACATCGGCAACTACAACCTGGGCGCGGCAACATCGGCGTCTACAACCTGGGCAG   | 9155 |
| QY  | 630  | la-----AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerA     | 647  |

Db 9156 CGCCAAACATCGGCGACTTCAACCTGGCGAGCGCCAAACATCGGCGACTTCAACCTGGCGAG 9215  
 Qy 647 lailethrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySers 667  
 Db 9216 CGCCAAACA-----TCGGCAGCTCC-AACATCGGCTTCGGCAACGTCGGTCCGGGG--- 9264  
 Qy 667 erTyrProAlaValPheValValGlyThrleGly---GlyValThrGlyAlaTyrArgs 686  
 Db 9265 -----CTCAGCGCGGCCCATCGCAACATCGGCTTCGGCAACACCGGAAACGGAACA 9316  
 Qy 686 erAspAspCysGlyThrThrTrpValLeuileAsnAspGlnHisGlnTyrGlyAsnT 706  
 Db 9317 TCGGCATCGGCAATACCGGACCGGCAACATCGGCTTCGGCAACACCGGAAACGGAACA 9376  
 Qy 706 rpGlyGlnAlaileThrGlyAsp--HisAlaAsnLeuArgValTyrile-GlyThrA 725  
 Db 9377 TCGGCATCGGCTGACCGCGGACACCATGACCGGCTTCGGCGGTGGAACATCGGCGACCG 9436  
 Qy 725 snGlyArgGlyIleVal-----TyrGlyAspIleGlyGlyAlaProSerGly 740  
 Db 9437 GCAACATCGGCGTATTCAACTCCGGCACCGCAACATCGGCTTCGGCAACCTCCGGC 9492

RESULT 2

US-11-052-554A-525  
 ; Sequence 525, Application US/11052554A  
 ; Publication No. US20050288866A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sachdeva, et al.  
 ; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
 ; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
 ; FILE REFERENCE: 30853/40359A  
 ; CURRENT APPLICATION NUMBER: US/11/052,554A  
 ; PRIOR FILING DATE: 2005-02-07  
 ; PRIOR APPLICATION NUMBER: US 60/589,227  
 ; PRIOR FILING DATE: 2004-07-20  
 ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
 ; PRIOR FILING DATE: 2004-02-06  
 ; NUMBER OF SEQ ID NOS: 763  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 525  
 ; LENGTH: 11151  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis H37Rv  
 US-11-052-554A-525

Alignment Scores:  
 Pred. NO.: 0.000763 Length: 11151  
 Score: 182.00 Matches: 187  
 Percent Similarity: 32.6% Conservative: 83  
 Best Local Similarity: 22.6% Mismatches: 325  
 Query Match: 4.5% Indels: 236  
 DB: 12 Gaps: 41

US-09-917-376-3 (1-740) x US-11-052-554A-525 (1-11151)

Qy 6 TyrThrTrpSerAsnValAlaileGlyGlyPheValAspGlyIleValPheAsn 25  
 Db 4190 TGGCGCTTGGGCAATCTCGGACGGGACGTCGGGTTCGGCAACATCGGTCGGCGCAACG 4249  
 Qy 26 GluGlyAlaProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTipAsp 45  
 Db 4250 TCGGGTTCGCCA-----ACTCGGTCGG----- 4273  
 Qy 46 AlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyr 65  
 Db 4274 -----CGGTGGGCC-----TGGCGCGCTCGGCAACGTCGGGT--- 4306  
 Qy 66 AsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaVal 85  
 Db 4307 -----TGAGCAATCGCGGACGCAACACTGGGGCTCGCCAAACCTGG 4348  
 Qy 86 GlyMetTyrThrAsnSerTrpAspProAsnAspGlyAlaileLeuArgSerSerAspGln 105

Db 4349 GTGTGGGCAACATCGGCTTGG---CACAACACCGGCA-----CGGGCAACATCG 4393  
 Qy 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArg 125  
 Db 4394 GGATCGGGCTGGTCCGGGACTACACAGCCGGCATCGGC-GGCCTCAAC-----TCG 4443  
 Qy 126 GlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAla 145  
 Db 4444 GGTAGTGGCAATATCGGATTGTTCATTCGGCACCGGCAATGTCGGGTCTTCTCAACACC 4503  
 Qy 146 ProSerGlyLys---GlyLeuTrpArgSer-----ThrAspSer 157  
 Db 4504 GGCACCGGCAACTTCGGACTGTTCAACTCCGGTAGTTTCAACACCGGCAATCGGTAATAGC 4563  
 Qy 158 GlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsnPro 177  
 Db 4564 GGAACCGGCACTACGGGCTCTTCATGCGGCAATTTCAACACCGGCAATCGGCAACCCC 4623  
 Qy 178 ThrAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTyrValAlaPheAspLys 197  
 Db 4624 GGG-----TCGTACAACACG 4638  
 Qy 198 SerSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn 217  
 Db 4639 GGCAGCTTCAATGTCGGT-----GATACCAAC 4665  
 Qy 218 AsnProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro 237  
 Db 4666 ACCGGTGGTTTCAACCCGGGCGACATCAACCCGGCTGG-----TTCAAC 4710  
 Qy 238 ThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257  
 Db 4711 ACCGGC-----ATTATGAATACGGGC 4731  
 Qy 258 ThrSerAsnThrGlyGlyProTyrAspGly-----SerSerGlyAspValTrpLysPhe 275  
 Db 4732 ACCCGCAACACCGGGCCCTCATGTCCGGGACCGACAGCAACGGCATGCTGGCGC--- 4788  
 Qy 276 SerValThrSerGlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsn 295  
 Db 4789 -----GGC 4791  
 Qy 296 AspTyr-----PheGlyTyrSer---GlyLeuThrIleAspArgGlnHisProAsn 311  
 Db 4792 GACCAGAGGGCGCTGTTCCGGCTGCTTATGTCATCAACGATC----- 4833  
 Qy 312 ThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThr 331  
 Db 4834 -----CCGCAATTCCCGATCCGATCCGATCACCACG 4860  
 Qy 332 AspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyrProAsnArgSerLeu 351  
 Db 4861 ACTGGCGTATCGGCCCCCATCGTCATCCGGACACACGATCCTTCCGCGCTGCACCTG 4920  
 Qy 352 ArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro 371  
 Db 4921 CAG-----ATCACCAGGACGCGGACTACAGCTTCACCGTG---CCGACATC 4965  
 Qy 372 ProValProSerProLysLeuGly-----TrpMetAsp 382  
 Db 4966 CCCATCCCGGCATCCACATCGGCATCAATGGCGTCGTCCACGTCGGCTTACCGGCCCCG 5025  
 Qy 383 GluAlaMetAlaileAspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAlaThr 402  
 Db 5026 GAAGCCACCTGCTGTCGCGCCCTGAAGAATAACGGTAGTCTTCATCAGCTTCGCGCCCATC 5085  
 Qy 403 LeuTyrAlaThrAsnAspLeuThrLysTrpAsp----- 413  
 Db 5086 ACGCTCTCGAATATCGAATATTCGCGCCCATGGATTTCAGCTTAGGGCTGCCCGTCTTGGT 5145  
 Qy 414 -----SerGly-----GlyGlnIleHisIleAlaProMetValLysGlyLeuGlu 429

Db 5146 CCTATACGGCGCAACTCGGACCAATTCATCTTGAGCCAATCGTGGT-GGCCGGGATCGG 5204  
Qy 430 ThrAlaValAsnAspLeuIleSerProSerProSerGlyAlaProLeuIle----- 445  
Db 5205 TGTGCCCTCGAGATCGAGCCATCCGCCCT---GGATCGCATTTCTGATGAGTGCAT 5261  
Qy 446 -----SerAlaLeuGlyAspLeuGlyPheThrHisAla 457  
Db 5262 TCCTATCCGCATACCTGTTGATATCCGGCTCGGTGTC-ATCGATGGATTTCAATGTCG 5319  
Qy 458 AspValThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSer 477  
Db 5320 GAAGTGTGCGATCGATGTCGTGGATCCCGGATCCCGGATCCAGTACAGGACCAACC 5379  
Qy 478 ValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer----- 493  
Db 5380 ATTTCCCGCATCCGCTGGCTTCGACATTCGCACCATCGCAGCATCGCCCAACATCCCG 5439  
Qy 494 ---PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGly 512  
Db 5440 ATCATCGACATCCCGCGCGCGCGCGGC-----TTCGGGAACATCGACCCAGATCCG 5490  
Qy 513 LysAsnTrpPheGlnGlySerGluProGlyValThrThrGlyValThrValAlaAla 532  
Db 5491 TCGTCGGGTTCTCAACACGGTCCGCGCGCGGATCGGCATCGGCAACTCGGTGCG 5550  
Qy 533 SerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyr 552  
Db 5551 GGCGTGCGGC-----CTGCTCAACAGCGCGCGCGGTCACGTGGGGACA 5601  
Qy 553 AlaValGlyPheGlyAsnSerTrpAlaAspGlnGlyValProAlaAsnAlaGlnIle 572  
Db 5602 CTCTCGGGGCTGGCAATGCGGCACCTCGCTCGGCTGTGCTGAACCTCGCGCACCGCC 5661  
Qy 573 ArgSer-----AspArgValAsnProLysThrPheTyrAla 584  
Db 5662 ATCTCCGGGCTGTTCAACGTGAGCACCTCGAGCCACACCGCGGGGTGATCTCGGG 5721  
Qy 585 LeuSerAsn-----GlyThrPheTyrArgSerThrAspGlyGlyVal 598  
Db 5722 TTCAGCAACCTCGGCGACCATATGTCGGGGTG-----TCCATCGATGGCTGATC 5772  
Qy 599 -----ThrPheGlnPro-----ValAlaAlaGlyLeu 607  
Db 5773 GCGATCCTCACTTCCACCTCGCGAGTCCGTGTTGATCAGATCATCGACGCGGCATC 5832  
Qy 608 -ProSer-----SerGlyAlaValGlyValMetPheHisAlaValProGlyLysG 624  
Db 5833 GCGAGCTCGAGCACCTCGACATCGGCAACGCTTTGGCCATGTGCGCGGGTG 5892  
Qy 624 uGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSe 644  
Db 5893 AACCTCGTTTGGTAACTGCTGAGTTCACCTGGGTGCGGCAACGTCGCGCAACATC 5952  
Qy 644 rTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaPr 664  
Db 5953 AACGTGCGGC---CCGCAACCTCGGCGGAGCAACTTGGGGTTGGCGACGTCGGGACC 6009  
Qy 664 o-----GlySerSerTyrProAlaValPheValValGlyThrIleGlyG 679  
Db 6010 GGCAACCTCGGGTTCGCGCAACATCGGTGCGGCAAT-TTCGGATTCGCGCAACGCGGCCT 6068  
Qy 679 yValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTrpValIleuIleAsnAsp 699  
Db 6069 GACCGCGCGCGCGGGGCTCGGCAATGCGGG-----FTGGGTAAACGCCGG 6116  
Qy 699 pGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeuArgAr 719  
Db 6117 CAGC-----GGCAGCTGGGGTTGGCCACGTCGGT---GTGGGCAATATCGGTT 6164  
Qy 719 gValTyrIleGlyThrAsnGlyArgGlyIle-----ValTyrGlyAsp----- 734  
Db 6165 GGCCAAACCGGACCGGCAACATCGGGATCGGGCTGACCGGGGACTATCGGACCGGGAT 6224

Qy 734 eGlyGlyAlaProSerGly 740  
Db 6225 CGCGCGCTGAACCTCGGC 6243  
RESULT 3  
US-11-136-527-2622  
; Sequence 2622, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wmth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 2622  
; LENGTH: 7231  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-2622

Alignment Scores:  
Pred. No.: 0.000712 Length: 7231  
Score: 180.00 Matches: 201  
Percent Similarity: 31.5% Conservative: 111  
Best Local Similarity: 20.3% Mismatches: 320  
Query Match: 4.5% Indels: 360  
DB: 12 Gaps: 49

US-09-917-376-3 (1-740) x US-11-136-527-2622 (1-7231)

Qy 40 GlyMetTyrArg-----TrpAspAlaAlaAsnGlyArgTrpIleProLeu--- 54  
Db 3121 GGAATATACAGAACTTCTGCTCTGGGACAGAAAT---CAGTGGACTTCTTCTCG 3177  
Qy 55 -----LeuAspTrpValGlyTrpAsnAsnTrpGly----- 64  
Db 3178 AGGAGACGATCAGAACTTCTACCTCTGGAATAGATGAGGACGCTGTCTTCTACTGG 3237  
Qy 64 ----- 64  
Db 3238 AAGAGAGGCTTAGAGAGCTGCTGCTGAGTAGAGACCTCAGTGGACTTCTTCTCG 3297  
Qy 65 -----Tyr-AsnGlyValValSerIleAlaAlaAspProIleAs 77  
Db 3298 AGAAGAGGTTCAAGAACTACTCTCTGGAATAGAGGACATAGTGTACTTCCGACTGG 3357  
Qy 77 nThrAsnLysValTrpAlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGl 97  
Db 3358 AGAAGTCCAGAACTTCTGCTCTGAGTGGGAGACTTGAGTGGACTTCCCTCAGGAGG 3417  
Qy 97 yAlaIleuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLeuProPhePh 117  
Db 3418 AGAAGTCTAGAAACATCTGCTTCAGGTGTAGAG---GATGTACCCAGCTTCTTACTGA 3474  
Qy 117 sLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAs 137  
Db 3475 AAGAGAGGCTTAGAGACTTCTGCTCTGGAATAGAGGACATCACTGTGTCTTCTACTGG 3534  
Qy 137 nAspAsnIleLeu-----TyrPheGlyAlaProSerG 148  
Db 3535 AAGAGAAATCTAGAACTTCTGCTCTGAGTAGAGGATGTCACTGTGACTTCTTCTCG 3594  
Qy 148 yLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnAspPr 168  
Db 3595 AAAAGAGGCTTAGAGACTTCTGCTCTGGAATA-----GAGGACATTAGTGTCTCC 3648  
Qy 168 oAsp-----ValGlyThrTyrIleAlaAsn---ProTh 178

|    |      |  |  |  |  |  |  |  |      |  |      |
|----|------|--|--|--|--|--|--|--|------|--|------|
| Db | 3649 | TACTGACGACGAGTCTAGAACTTCTGCGCTCCGGGGATATGTCAGTGGGATTCCTTC        |  |  |  |  |  |  | 4650 | GGATTTCTGATCTCAGTGGGCAACCTTCAGGGTTCCAGTGTCTCAGGGACACACCTCTG    | 4709 |
| Qy | 178  | rAspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSe     |  |  |  |  |  |  | 449  | -----GlyAspLeuGlyGlyPheThrH                                    | 456  |
| Db | 3709 | TGGAGAGATGGTACAGAACTCTACTTCTGGAGTAGAGGGTGTGAGT-----GG            |  |  |  |  |  |  | 4710 | GAACCCCTGACCTGGCTTCTGGCGCATGAGTGGCAGTGGAGATTTCTTCTGGCATTACGT   | 4769 |
| Qy | 198  | rSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAs     |  |  |  |  |  |  | 456  | isAlaAspValThrAlaVal-----ProSerThr                             | 465  |
| Db | 3760 | TCTTCTCTGGAGAGAGGGTCTAGAAACGCTCTGCCCTCTGGAGTAGAAGATCTTGGTCT      |  |  |  |  |  |  | 4770 | TTGTGGACACCACTCTATTGAAGTGACCCCTACCACTTTAGAGAAGAAGAGGGGTTAG     | 4829 |
| Qy | 218  | nProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProTh     |  |  |  |  |  |  | 465  | -----  | 465  |
| Db | 3820 | TCCTACAGATAGTCTAGAGACTTCTGCCCTCAGGAGTAGATGTACTGGGTATCCCTTC       |  |  |  |  |  |  | 4830 | GGTCTGTGAACTCAGTGGCCCTCCCTTCTGGGGAGACAGATCTGTCTGGCACATCTGGGA   | 4889 |
| Qy | 238  | rGly-----PheIleProHisLysGlyValPheAspProValAsnHi                  |  |  |  |  |  |  | 466  | -----IlePheThrS  | 469  |
| Db | 3880 | TGGAGAGAGACACAGAACTCTGTCTCT-----GGGTAGGTAGTACCTTAGTGG            |  |  |  |  |  |  | 4890 | TGGTGGATGTCTCAGTGGACAACTCTTCTGGAGCAATCGATTCCAGTGGACTATATCCCCA  | 4949 |
| Qy | 252  | sValLeuTyr-----IleAlaThrSerAsnThrGly-----                        |  |  |  |  |  |  | 469  | erProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIle       | 487  |
| Db | 3934 | ACTTCTCTGGCAAGAGGCTTAGAGACCTCAGCTTCAGGAGCTGAGGACCTTGGTGG         |  |  |  |  |  |  | 4950 | CTCCAGAAATTC---AGTGGCTCCCAAGCGGAGTAGCTGAGGTCACTGGAGAGGTCTCTG   | 5006 |
| Qy | 263  | -----GlyProTyrAsp-----GlySerSerGlyAspValTrpLysPhe-----           |  |  |  |  |  |  | 498  | -----IleValArgAlaGlySerPheAsp                                  | 495  |
| Db | 3994 | CTTGCTTCTGGAAGAGAGACTTGGTAGGGTCTGCTTCTGGAGCCCTGGACTTCGGCAA       |  |  |  |  |  |  | 5007 | GAGTTGAGACTGGGAGACGCTTGTCTCGGGAGCATTCGATGCCAGTGGACTTGTCTCAG    | 5066 |
| Qy | 276  | -----SerValThrSerGlyThrTrpThrArgIleSerProValProSerTh             |  |  |  |  |  |  | 496  | -----ProSerSerGlnProAsnAspArgHisValAlaPheSer                   | 508  |
| Db | 4054 | GCTACCTTCTGGAACCTCTAGGAAGTGGTCAAACTCCAGAACTAGTGGCCCTTCCCTCTG     |  |  |  |  |  |  | 5067 | GTTTCCCTACGGTGTCTCTGTAGACAGAACTTTGGTGGAGTCTATAACCTCGCTCCTA     | 5126 |
| Qy | 291  | rAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeu-ThrIleAspArgGlnHisProA     |  |  |  |  |  |  | 509  | -----ThrAspGlyGlyLysAsn-----TrpPheGlnGlySerGluProG             | 522  |
| Db | 4114 | ATTAGTGGTGAA-----TATTCTGGAGTGACATGGAAGTGGCCCATCTC                |  |  |  |  |  |  | 5127 | CTGCTCAAGAACTGGAGAGGGCCCTTCAGCATTTTGGAAATTCAGTGGTGCCCATCTCTG   | 5186 |
| Qy | 311  | snThrIleMetVal-----AlaThrGlnIleSer-----                          |  |  |  |  |  |  | 522  | lyGlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrpA      | 542  |
| Db | 4162 | TGGCGTGGCTGACTTGTAGTGACTTCCATCCGGCTTCCCAACAGTCTCCCTTGTGGACAG     |  |  |  |  |  |  | 5187 | GAACACACAGACATATCTGGGACCTCTCTGGGTCTTTAGACCAAGC-----ACATGGC     | 5240 |
| Qy | 321  | -----Trp-----TrpProAsp-----                                      |  |  |  |  |  |  | 542  | laProGly-----AspProGlyGlnProValValTrpAlaValGlyP                | 556  |
| Db | 4222 | TACCTTAGTGAAGTATCAGACCCACCATCCAGTGAACCTGGAAGGAGGGGACCAT          |  |  |  |  |  |  | 5241 | AGCTTGGGTGGAGAGAGCCAGACAGAGCCACCAAGCTCCCATATTTTAGTGGGAGCT      | 5300 |
| Qy | 325  | -----ThrIlePheArgSerThrAspGly-----GlyAlaThrTrpThrArgIleT         |  |  |  |  |  |  | 556  | heGlyAsnSerTrpAlaAlaSer-----                                   | 563  |
| Db | 4282 | CAGCGTAGTGGTTCAGGAGAGAGTCAAGGCGGCCCTCAGTGAAGTGGACAGTAGTGC        |  |  |  |  |  |  | 5301 | TCTTAGCACCACTGATGCAAGTGGAGATCCATACTGCCCACTGGCAGTGGGAAA         | 5360 |
| Qy | 341  | rpAspTrpThrSerTyrProAsnArgSer-----LeuArgTyrValLeuA               |  |  |  |  |  |  | 564  | --GlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal-----              | 577  |
| Db | 4342 | G-GACATTAGTGGTCTCCCTTCAGGAACCTGAACCTCAGTGGGCCAACATCCGGATCTCTG    |  |  |  |  |  |  | 5361 | CCTCGGGGCTTCCGGAAGTCACTTTAATCACTTCAGAGTTAGTGGAGGGGTGACTGAAC    | 5420 |
| Qy | 356  | spIleSerAlaGluProTrpLeuThrPheGlyValGlnProAsnPro-----             |  |  |  |  |  |  | 578  | -----AsnProLysThrPheT  | 583  |
| Db | 4401 | ATGTCAGTGGAGAAACATCTCGAATTTTGTATGTAGTGGACAGCAATTTGGTCTTCTG       |  |  |  |  |  |  | 5421 | CAACTGTTTCCAGGAACCTTGGCCATGGTCTTCTTATGACATATCTCCCGGCTCTTTG     | 5480 |
| Qy | 372  | -----ProValProSerProL  |  |  |  |  |  |  | 583  | yrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyValThrPheGlnPro-   | 602  |
| Db | 4461 | GCACGTGGTGGGAACATCTCGGATTTCTGAGTCTCTGAGTCAGTGGACAGCCCTCAGGAGTCTG |  |  |  |  |  |  | 5481 | AGGCCAGTGGGGAAGCCCTCAGCATCTGGGGACCTTGGTGGACCTGTAAACAATCTTTCCCG | 5540 |
| Qy | 377  | ysLeuGlyTrpMetAspGluAlaMetAlaIleAspProPheAsnSerAsp-----          |  |  |  |  |  |  | 603  | -----ValAlaAlaGlyLeuProSerSerGlyAla                            | 613  |
| Db | 4521 | ACACCAGTATATCTGAGCTTAGTGACCTCTCTGGGCAACCAAGATGTCTAGTGGAG         |  |  |  |  |  |  | 5541 | GGTCTGGGTAGAACTTCAGTCCCAAGAGGAGCAGTACCCTCTGCTTACCCTGAGG        | 5600 |
| Qy | 394  | -----ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrL        |  |  |  |  |  |  | 613  | alGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerS   | 633  |
| Db | 4581 | AAGGCTCTGGAATCTCTTTGGCAGTGGCCCAATCCTCTGGCATAACATCT-----          |  |  |  |  |  |  | 5601 | CTGAGTGGGAGTGTCTGCTGCCCT-----GAGGCCAGCA                        | 5636 |
| Qy | 411  | ysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrA        |  |  |  |  |  |  | 633  | erGlyLeuTyrHisSerThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerS   | 653  |
| Db | 4631 | -----GTGAGTGA-----GAAACCTCTG                                     |  |  |  |  |  |  | 5637 | GTCAATTTG-----TCTGAGTCCAGATCTGCATGTGAATCACCT                   | 5675 |
| Qy | 431  | laValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeu-----       |  |  |  |  |  |  | 653  | erAlaVal---AsnValGlyPheGlyLysSerAlaProGly-----SerSerTyrP       | 669  |
|    |      |  |  |  |  |  |  |  | 5676 | CTGCTCCCGTGAACAGATCTGGAGATGCACAAACCCAGGCGGAGGTGAGGACCAACC      | 5735 |

|   |         |  |          |
|---|---------|--|----------|
| Alignment Scores:                                     |         |  |          |
| Pred. No.:  | 0.00133 | Length:  | 9474     |
| Score:  | 178.00  | Matches:   | 191      |
| Percent Similarity:                                   | 33.4%   | Conservative:  | 113      |
| Best Local Similarity:                                | 21.0%   | Mismatches:  | 296      |
| Query Match:  | 4.4%    | Indels:  | 317      |
| DB:   | 12      | Gaps:  | 47       |
| US-09-917-376-3 (1-740) x US-11-052-554A-526 (1-9474) |         |  |          |
| QY  | 43      | ArgTrpAspAlaAlaAsnGlyArg                                     | ----- 50 |
|   |         |  |          |
|   |         | :::  | :::      |
| Db  | 521     | CGTTGGGGGTGGTGTGCGCGCGTCTCAATGCTCTATTTCGGGACCGCGGAGATGT      | 580      |
| QY  | 51      | -----Trp---lIleProLeuLeuAspTrpValGlyTrpAsnTrp                | 63       |
|   |         |  |          |
|   |         | :::  | :::      |
| Db  | 581     | TCAGGCTTAAACCGCGGTCTGGGCAATGTGCGTAATTACAACGTCGGGTGGGCAATGTGC | 640      |
| QY  | 64      | GlyTyrAsnGlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrp    | 82       |
|   |         |  |          |
|   |         | -----TCAACTGGCGGAGCGCAATGTCGTGGCGCAGAAATTTGGGG               | 688      |
| Db  | 641     | GGATAT-----  |          |
| QY  | 83      | -----AlaAlaValGlyMetTyrThrAsnSerTrpAspProAsnAspGly           | 97       |
|   |         |  |          |
|   |         | :::  | :::      |
| Db  | 689     | CTGCCAACCGCGGTAGCGGNAATTCGGTTTCGGCAATATCGGCAACGCCAATTCGGGT   | 748      |
| QY  | 98      | AlaIleLeuArg-----  | 101      |
|   |         |  |          |
|   |         | :::  | :::      |
| Db  | 749     | TCGGCAACTCGGCTTTGGGTTGCCCGCGGCATGGGCAATATTGGTTGGGCAATGCGG    | 808      |
| QY  | 102     | -----SerSerAspGlnGlyAlaThrTyrGlnIleThrProLeuProPhelLys       | 117      |
|   |         |  |          |
|   |         | -----CGAGCAACTCGGAGCTCGCAACCTGGGTGTGGGCAACATCGGTTCGTCACACGG  | 868      |
| Db  | 809     | CGAGCAACTCGGAGCTCGCAACCTGGGTGTGGGCAACATCGGTTCGTCACACGG       |          |

QY 361 -----ProTrpLeu 363  
Db 1906 ATCAGATCGACCGCATCCCGTGAACCTCGCGCGCCAGCGTCACTGTGCGCCCTATCCTG 1965  
QY 364 ThrPheGlyValGlnProAsnProValProSerProLysLeuGlyTrpMetAspGlu 383  
Db 1966 ATCAACGGGGT-----AATATCCGGCG--ACCCCGGGCTTTGGCAACACGACCACC 2016  
QY 384 AlaMetAlaLeuAspProPheAsnSer---AspArgMetLeu-----Tyr 397  
Db 2017 GTCGCGTGTGGGTCTTCTCAACTCCGGCAGCGTGGGTGTGCGGGTCTCGGGAATTC 2076  
QY 398 GlyThrGlyAlaThrLeuTyr-----AlaThrAsnAspLeuThrLysTrpAspSer 414  
Db 2077 GGTGGGCGACCTCGGGTGTGGGAACACGCGACCGAGGTGGCTGGGGCGGGTTCG 2136  
QY 415 Gly-----GlyGlnIleHisIleAlaProMetValLys 425  
Db 2137 GGTTCGCCAATTTCCGTTCGCTGGGATCGGTGTGCTGAACCTCGGCTCGGGTGTGTCG 2196  
QY 426 GlyLeuGluGluThrAlaValAsnAspLeuIleSerProSerGlyAlaProLeuIle 445  
Db 2197 GGGCTGTACACACCGCGCGG-----TTGCCCGCGGGAGCCCGCGGTGTC 2244  
QY 446 SerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThr 465  
Db 2245 TCGGCATCGGCAATGTTGGTCAG----- 2268  
QY 466 IlePheThrSerProValPheThrThrGlyThrSerValAspTyrAlaGluLeuAsnPro 485  
Db 2269 ---CAGCTGTGGGGTGTCTCGCGCGGGAGCGCA-----CTCAACACG 2310  
QY 486 SerIleIleValArgAlaGlySerPheAspProSerGlnProAsnAspArgHisVal 505  
Db 2311 AGCCTCATCAATCTCGGGTGGCCGATGTGGCAGCGTA-----AACGTC 2358  
QY 506 AlaPhe-SerThrAspGlyGlyLysAsnTrpPheGln----- 517  
Db 2359 GGTTCGGCAACGTTCGGGACTTCAACCTGGGTGGCGCAATATCGCGACTTGAACGTG 2418  
QY 518 ---GlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAla-----SerAl 534  
Db 2419 GGTTCGGCAATGTTCGGCGCGGCAACCTCGGGTTCGGCAATATCGCGGATGCCAATTC 2478  
QY 534 aAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVal 554  
Db 2479 GGGTGGCAATGCGGGTCTGGCGCGGGCTCGCGCGG----- 2518  
QY 554 lGlyPheGlyAsnSerTrpAlaAlaSer---GlnGlyValProAlaAsnAlaGlnIleAr 573  
Db 2519 -----TGGGCACATCGGGTTCGGCAATGCGGAGCGGCA----- 2554  
QY 573 gSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSe 593  
Db 2555 -----ACGTCGGCTTCGGCAACATGCGGTGGGC 2583  
QY 593 rThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeu----- 607  
Db 2584 AACATCGGGTTCGGTAACACCGGCAACCAACCTCGGGATTCGGCTGACCGGGGACAC 2643  
QY 608 -----ProSerSerGlyAlaValGlyValMetPheHisAl 619  
Db 2644 CAGACTGGGATCGGGCTTGAACTCCCGT--GCCGGCAACATCGGG-----TTGTTCAATC 2699  
QY 619 aValProGlyLysGluGlyAspLeuTrpLeuAlaAlaSerSer-----GlyLeuTyrHi 637  
Db 2700 CGGCACCGGCAACGTCGGG---TTGTTCACTCCGGGACCGGGAATTCGGGTTCGTTCAA 2756  
QY 637 sSer-----ThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSe 653  
Db 2757 CTCGGGCAAGCTTCAACACCGGATCGGCAATGCGGGAACGGGCGAGTACTGGGCTTTCAA 2816  
QY 653 rAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheVa 673

Db 2817 TGCCGGTAATTTCAATACCGGTGTGCCCAACCTTGGTGTGTCACACACGCGGCTTCAA 2876  
QY 673 lValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspAspCysGlyThrThrTr 693  
Db 2877 TGTGGGTGACACCAAC-----ACCGTGTGTTTCAACCCGGGCGAGCATCAACACCGGCTG 2930  
QY 693 pValleuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAs 713  
Db 2931 GTTCAACACCGGCAACGC-CAACACCGGCTGCGCAATTCGGGCAA----- 2975  
QY 713 phisAlaAsnLeuArgValTyrIleGly 723  
Db 2976 ----TGTGCACACCGCGGCTCATGTGCGG 3002  
RESULT 5  
US-11-052-554A-527  
; Sequence 527, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 527  
; LENGTH: 7572  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis H37Rv  
US-11-052-554A-527  
Alignment Scores:  
Pred. No.: 0.00166 Length: 7572  
Score: 175.50 Matches: 175  
Percent Similarity: 33.3% Conservative: 92  
Best Local Similarity: 21.8% Mismatches: 327  
Query Match: 4.3% Indels: 209  
DB: 12 Gaps: 38  
US-09-917-376-3 (1-740) x US-11-052-554A-527 (1-7572)  
QY 10 AsnValAlaIleGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAla--- 28  
Db 4564 AATACCGGATCGTAAATTCGGGGATTGCCAGCACCGGGTGTTCACCGGGTGGGTTTC 4623  
QY 29 ----ProGlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrpAspAlaAla 47  
Db 4624 AACACCGGTGTGTCATCGGTAGCTACACACCGCGAGTTTCAACCGCGGCGAGGCC 4683  
QY 48 Asn---GlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsn 66  
Db 4684 AATACCGGCGGTTCACACCGGCGAGTGTCAACACCGGCTGGTTGAACACCGGT----- 4737  
QY 67 GlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGly 86  
Db 4738 -----GACATCAACACCGGG----- 4752  
QY 87 MetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSer---AspGln 105  
Db 4753 ---GTGGCCAACTCGGCGAGCTCAACACCGGCGCTTCATCTCCGTAACAGCAAC 4809  
QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGly--- 124  
Db 4810 GCGCGCTTCGCGCG-----GGCAGTACTACCGGCGCTG 4842  
QY 125 ArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGly 144





|    |       |  |       |
|----|-------|--|-------|
| Qy | 27    | GlyAlaProGlyIleLeuTyrValArgThrAspIleGlyMetYrArgTrpAspAla     | 46    |
| Db | 12556 | GGCGCTAATAGTTCGCGAATAATCAGTACCGACAACGGCGCA                   | 12597 |
| Qy | 47    | AlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsn | 66    |
| Db | 12598 | -----ACCTGGGTGAAC-----                                       | 12609 |
| Qy | 67    | GlyValValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGly | 86    |
| Db | 12610 | -----GTGACCGTCGCGCAGACAGCCTGAC-----                          | 12636 |
| Qy | 87    | MetTyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGly | 106   |
| Db | 12637 | -----TGGAGTTACGTTGACGCGACGAACCTCACCAACGGCACC-----            | 12675 |
| Qy | 107   | AlaThrTrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGly | 126   |
| Db | 12676 | ACCACCTGCAGGTGCGGTGGTC-----GATCTGGCGGGCAACGTT-----GGCGCAACG  | 12726 |
| Qy | 127   | MetGlyGluArgLeuAlaValAspProAsnAsn-----                       | 137   |
| Db | 12727 | AGCAGCCAGTCGGCGTGATCGATACACGTTAAACCCGCGCGAGTGCTCACCATGCCACG  | 12786 |
| Qy | 138   | -----AspAsnIleLeuTyr-----                                    | 142   |
| Db | 12787 | ATCAGACCCGACACGGGGAGTTCGGCACTGACTTTATCACCAGCGCACCATGCTCAG    | 12846 |
| Qy | 143   | Phe-----GlyAlaProSerGlyLysGlyLeuTrpArgSerThrAsp              | 156   |
| Db | 12847 | CTGACCGGTTGCTGGGGCGCGGCTTGCCAGCGCGAAGTGGCGCAGATTAGCCTTCAT    | 12906 |
| Qy | 157   | SerGlyAlaThrTrpSerGlnMetThrAsnPheProAspValGlyThrTyrIleAlaAsn | 176   |
| Db | 12907 | AGCGCGCACTTGGACCAACGCTGCCACCAACCGGTACACAGTGACATTACACCGACAGC  | 12966 |
| Qy | 177   | ProThrAspThrThrGly-----TyrGlnSerAspIleGlnGlyValValTrpVal     | 193   |
| Db | 12967 | CGCACGCTCACCAGCGGAGCTACGTTTATCAGTGGCGGGTG-----               | 13008 |
| Qy | 194   | AlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal    | 213   |
| Db | 13009 | ---CTGGATCTGGCGGGAAACACCGCCCGGTGGTGTGAAACCGTGGTGTGCATACG     | 13065 |
| Qy | 214   | AlaAspProAsn-----AsnProValPheTrpSerArgAspGlyGlyAlaThrTrp     | 230   |
| Db | 13066 | ATTAAACCCACCGCCACACCAACGATGTGTGCTATCGATGATGTCTGGCAGCGGCGAG   | 13125 |

```
QY 471 Val---PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleVal 489
Db 14008 ATTCAGTGCACCGTCGCGCACCTGACGTGGTACGCGC-----TCGATTGTG--- 14052
QY 490 ArgAlaGlySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThr 509
Db 14053 -----GCATTGTATAAGAGGCG----- 14070
QY 510 AspGlyGlyLysAsnTrpPheGlnGlySerGluProGly----- 522
Db 14071 GACGGCTATCTCGACTTCTGGATTGGTGCAGCTGCGCGCGGAGCTCCACACCTTCCTCG 14130
QY 523 -----GlyValThrGlyGlyThrValAlaAlaSerAlaAspGlySerArg 538
Db 14131 TGGAAACAGCAGCAGCAGCTGGTAGGCAACTCCACCGTCGACACAGCGCGGTAGCGCC 14190
QY 539 PheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsn 558
Db 14191 ACGTGGCGCGG-----GCGGTGACGGGTATCTT 14220
QY 559 SerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgVal--- 577
Db 14221 TCGCTCAACGAAGTTCGTGGCGTCGATCTGACAAATGACGCGAGGATCGACTGTTTCAG 14280
QY 578 -----AsnProLysThrPheTyrAlaLeuSer-----AsnGly 588
Db 14281 CACACCTATACCTCAACCACTATTACACGCTGCTTCGCTCATCAACACGAGGAATGG 14340
QY 589 ThrPheTyrArgSerThrAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuPro 608
Db 14341 ACGTTTGTCTGGGCGAGAACACCAACCAATACCTTCCTGACGCGGCGGCGAGTGGCGCT 14400
QY 609 SerSerGlyAlaValGlyValMetPheHisAlaValProGlyLys---GluGlyAspLeu 627
Db 14401 ATGAGCAGCAGCGTTTCCATACCTACCTGGGCCGATTCGATGTGTCGAGCGCATATGATCTC 14460
QY 628 TrpLeuAlaAlaSerSerGly-----LeuTyrHisSerThrAsnGly 641
Db 14461 TTCCTGCCCGCGCAGCAGGGAAGAGCTAACTACGCTCGCTGTTATTCAACACCAACGCG 14520
QY 641 ----- 641
Db 14521 GTACTGGGTTCGCGGTGGCGGTGGGCGCAACCGCAACCACTACGCCAGCCAGTTTACG 14580
QY 642 -----Gly 642
Db 14581 CTGGCGGTGGACTGGAACCAACGAGCGGCTGATGGATATCGCCCGTATGCCCGCAGACCGGG 14640
QY 643 SerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 14641 CAGTCGTAT---CTTTATACTAAGTCAAGTCAGCAACGCCAGCAAC-----TGGACGCAATCG 14691
QY 663 AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThr--- 681
Db 14692 GCCCTCGCGCGCAGCCAGAGC-----GTTACCAACGAGCGCGGTGGCGCA 14736
QY 682 -----GlyAlaTyrArg-----SerAspAspCysGlyThrThr 692
Db 14737 ATGACTACGACTGGAGCGCGCGGTGGATGTGTGTGTCCAAACAGTCGCGGAGCGGTG 14796
QY 693 TrpValLeuIleAsnAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGly 712
Db 14797 TTCCTGAGCGCCACACCAACCAACCGGTGAGTACGGC-----ACTCG 14838
QY 713 AspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArgGlyIleValTyrGly 732
Db 14839 CTACACCTCGGCATC-----ACCGATCCCAACGCAATTAAGTCTATTACGGC 14886
QY 733 Asp 733
Db 14887 AAT 14889
RESULT 7
```

```
US-11-052-554A-401
; Sequence 401, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 401
; LENGTH: 2943
; TYPE: DNA
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-401
```

```
Alignment Scores:
Pred. No.: 0.0026 Length: 2943
Score: 166.50 Matches: 164
Percent Similarity: 31.2% Conservative: 117
Best Local Similarity: 18.2% Mismatches: 302
Query Match: 4.1% Indels: 319
DB: 12 Gaps: 42
```

US-09-917-376-3 (1-740) x US-11-052-554A-401 (1-2943)

```
QY 57 TrpValGlyTrpAsnAsnTrpGly----- 64
Db 169 TATGTAGCTGAACAACTATGCTCCCTGATGATATTACAGCGGATATCTACCGGTC 228
QY 65 -----TyrAsnGlyValValSer-----IleAlaAlaAspProIleAsn 77
Db 229 ATTAATAACGCTACCGTTAACGCGGTGATTCTACTACTACTACTACTACTACTACTACTACT 288
QY 78 ThrAsn-----LysValTrpAlaAlaValGlyMetTyrThr 89
Db 289 ACCAATACCAACGCAATAGTCTGACAATCAAAAACAGCACTATTTCACGGTATGATTACC 348
QY 90 AsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrp 109
Db 349 TCTGAGTCATGACTACTGATTGTGCT-----GATGACCGTCTACTGTT 393
QY 110 GlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGlu 129
Db 394 TATGTTAT-----GAT 405
QY 130 ArgLeuAlaValAspProAsnAsn-----AspAsnIleLeuTyrPheGlyAla 145
Db 406 CGTCTGACACTGAGCGTTGATAATTCAACGATCGATGACAACTACGAGCATTA 459
QY 146 ProSerGlyLysGlyLeuTyrArgSerThrAspSerGlyAlaThrTrpSerGlnMetThr 165
Db 460 -----ACTTACAACGGTACTATAATAATGCCGCTGACACTCATGTTGTA 504
QY 166 AsnPheProAspValGlyThr----- 172
Db 505 GATGTTTACGATATGCGGTACTGCTATTACCTGGATCAGGAAGTTGATGTGCCATCACT 564
QY 173 -----TyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGlnGly 189
Db 565 AATAACTCTCATGTAGCAGGTATTACGCTGACTCAGGGTTATGAG----- 609
QY 190 ValValTrpValAlaPheAspLysSerSerSerLeuGlyGlnAlaSerLysThrIle 209
Db 610 -----TGGGAAGATATTGACGACAAACACAGTCAGCACTGGCGTAAACAGCAGCAAGTG 663
QY 210 PheValGlyValAlaAspProAsnAsnProVal----- 220
```

Db 664 TTT-----AATAACACTTATTACTGTTAAAGATTCTACTGTGACCTCT 705  
Qy 221 ---PheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
Db 706 GGTTCATGAGTGAAGGT-----ACTACTGGT 735  
Qy 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTyr-----Ile 256  
Db 736 TGGTTTGCCATACTGTTAAATGCCAGCAACTATAGCAACAGCTGACTGCAGACGATGT 795  
Qy 257 AlaThrSerAsnThrGlyGlyProTyrAspGlySerSerGlyAspValTrpLysPheSer 276  
Db 796 GCAATTCCGCAATCGCAAAATCCGTAT-----GCTGATAATGCGATCGACGACTACA 846  
Qy 277 ValThrSerGlyThrTrpThrArgLysSerProValProSerThrAspThrAlaAsnAsp 296  
Db 847 GTAACCTTAGCAACTACACTGATGGTGATGTTGTTTCTCCAGTAATTTGATGAA 906  
Qy 297 TyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAla 316  
Db 907 AACTTCTCCGCAAGTGTAAACAGCTATCGCGATGCTGATGGTGATGA----- 957  
Qy 317 ThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThr---AspGlyGlyAla 335  
Db 958 ---GATACCAACCGTTGGGTGGACACAGCCTATGGATGTCACTCTGAACAACGGCAGC 1014  
Qy 336 ThrTrpThrArgIle-----Trp 341  
Db 1015 AAGTGGTGGCGCTGCAATGCTGTTTCATATGTTGATGAAGATGGTGATGTTCTTAC 1074  
Qy 342 AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro 361  
Db 1075 GACGATATGCTGTGGTACTGAAGCAACTGCACTCTGCTGATATTCGACTTAACAGC 1134  
Qy 362 TrpLeuThrPheGlyValGlnProAsnProProValProSerProLysLeuGlyTrpMet 381  
Db 1135 CTG-----TGGCCT 1143  
Qy 382 AspGluAlaMetAlaIleAspProPheAsnSerAsp----- 393  
Db 1144 TCATCAACTGTCGGTGTGTGAATCAATCAATCAATCAATATGACGAAATGGCCATATCGTA 1203  
Qy 394 ---ArgMetLeuTyrGlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrp 412  
Db 1204 GGAACGAGTTTACGAGCGGTTGTTTAAATGTGACTTTTGAACGGTGGTTCAGAGTG 1263  
Qy 413 Asp-----SerGlyGlyGlnIleHis 419  
Db 1264 GATACACAAAATCTCTCTGATTGATACTTTAAGTATTACAGCGGTTCCCAAGTTAAT 1323  
Qy 420 IleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProPro 439  
Db 1324 GTTGCA-----GACTCTGCTGTGATCTCTGACACTGTCTCTGACT 1365  
Qy 440 SerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspVal 459  
Db 1366 GCGCGTTCTAACCTG-----AACATCGGTGAAGACGGTCTATGAGCGACTAATACCCCTG 1419  
Qy 460 ThrAlaValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAsp 479  
Db 1420 ACCATCGACATAGTACCGTTAAATGCTGATGATGTTTCTGCGGCTGGGGTTAGAA 1479  
Qy 480 TyrAlaGluLeuAsnProSerIleIle-----ValArg 490  
Db 1480 GATGCTGACTGTACGCAATACCATCACCGCTAACTACAGCGTCTGTTGATATTAAAC 1539  
Qy 491 AlaGlySerPheAspProSerSer---GlnProAsnAspArgHisValAlaPheSerThr 509  
Db 1540 GTTGATCAGTTTCGATGTCAACCGTTCCAGGCCGATACCTGGAATCTGACCACTACCACT 1599  
Qy 510 AspGlyGlyLysAsnTrpPheGlnGly----- 518

RESULT 8

US-11-052-554A-518

; Sequence 518, Application US/11052554A

Db 1600 GATACTAACGGCAACATTCAACGCTGGTGTATTTCGATATCCATAGCAGTGATTACGTAATG 1659  
Qy 519 -----SerGluProGlyGlyValThrThrGly-----Gly 528  
Db 1660 GATACCGATCTGGTCAACGATCGTACCAACGATACCTACCAAGTCAAACTACGGTTATGGC 1719  
Qy 529 ThrValAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGln 548  
Db 1720 TTAATCGCAATGAACCTCTGATGCTACCTGACTATTAAACGGTAAACGGCGATAACGACAAC 1779  
Qy 549 ProValValTyrAlaValGly-----PheGlyAsnSerTrpAlaAla 562  
Db 1780 ACTGCTTCTATCGAAGCTGGTCAGAACGAAGTTGATAACCAACGCTGACCATGTTGACGCC 1839  
Qy 563 SerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArg----- 576  
Db 1840 GCGACCGGT-----AACTACAAAGTTGTTATCGACAACGCTACTGCTGCTGTTCT 1890  
Qy 577 -----ValAsnProLysThrPheTyrAlaLeu 585  
Db 1891 ATCGCTGACTACAACGGCAACGAGCTGCTACGTCAACGACAAAC----- 1938  
Qy 586 SerAsnGlyThrPhe-----TyrArgSer 593  
Db 1939 AGCAACGCGACCTTCTGCTGCTAAACAAAGCTGACCTGGTGCATACACTATCAGGCT 1998  
Qy 594 ThrAspGlyGlyValThr-----PheGlnProVal 603  
Db 1999 GAACGCGCGGTAAACACCGTTGTTGCAACAGATGGATGACCGACTACGCTAACATG 2058  
Qy 604 AlaAlaGlyLeuProSerSerGly-----AlaValGly 614  
Db 2059 GCGCTGAGCATCCCATCTCGCAACCAATATCTGGAACCTCGAACAAGACACCGTTGGT 2118  
Qy 615 ValMetPhe-----HisAlaValProGlyLysGlyAspLeuTrpLeuAla 630  
Db 2119 ACTCGTTTGACCAACTCTGCTCATGCGCTG---GCTGATAACGCGCGCATGGTAAAG 2175  
Qy 631 AlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAla---IleThr 649  
Db 2176 TACTTCGGTGGTAACTTCAACGCGCAACCGGCAACCATCAACTATGATCAGGATGTTAAC 2235  
Qy 650 GlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSerTrpPro 669  
Db 2236 GGCATCATG-----GTGCGTGTGATACCAAAATTGACGCGTAAAC----- 2277  
Qy 670 AlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAsp----- 687  
Db 2278 GCTAAGTGGATCGTCGGTGGCTGCAGGCTTCGCTAAAGGTGACATGAATGACCGTTCT 2337  
Qy 688 -----AspCysGlyThrThrTrpVal----- 694  
Db 2338 GGTCAAGTGGATCAAGACAGCCAGACTGCTACTACTCTTCTGCTCATTTCGCGAAC 2397  
Qy 695 ---LeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAsp 713  
Db 2398 AACGCTTTGTTGATGTAGTTGAGCTACTCTCACTTCAACACGACCTGCTCGAACCC 2457  
Qy 714 HisAlaAsn-----Leu 717  
Db 2458 ATGACCAACGGTACTTACGTTGACGGTAGCACCAACTCCGACGCTTGGGGCTTCGGTTG 2517  
Qy 718 ArgArgValTyr-----IleGlyThrAsnGlyArgGlyIleValTyrGlyAspIle 734  
Db 2518 AAAGCGGTTACGACTTCAAACTGGGTTGATGCTGGTTACGTGACTCCTTACCGCAGCAT 2577  
Qy 735 GlyGly 736  
Db 2578 TCTGGT 2583

Publication No. US20050288866A1

# GENERAL INFORMATION:

APPLICANT: Sachdeva, et al.  
 TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
 TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

FILE REFERENCE: 30853/40359A

CURRENT APPLICATION NUMBER: US/11/052,554A

CURRENT FILING DATE: 2005-02-07

PRIOR APPLICATION NUMBER: US 60/589,227

PRIOR FILING DATE: 2004-07-20

PRIOR APPLICATION NUMBER: IN 173/DEL/2004

PRIOR FILING DATE: 2004-02-06

NUMBER OF SEQ ID NOS: 763

SOFTWARE: Patentin version 3.3

SEQ ID NO 518

LENGTH: 6615

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis H37Rv

US-11-052-554A-518

## Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00946 | Length:       | 6615 |
| Score:                 | 165.00  | Matches:      | 196  |
| Percent Similarity:    | 30.8%   | Conservative: | 88   |
| Best Local Similarity: | 21.3%   | Mismatches:   | 292  |
| Query Match:           | 4.1%    | Indels:       | 348  |
| DB:                    | 12      | Gaps:         | 45   |

US-09-917-376-3 (1-740) x US-11-052-554A-518 (1-6615)

|    |     |  |     |
|----|-----|--|-----|
| Qy | 7   | ThrTrpSerAsnValAlaLeuGlyGlyGlyPheValAspGlyLeuValPheAsnGlu    | 26  |
| Db | 5   | ACCTGGTCTCC-ACAACGTCGGGAATGTCGGGCTTCTC                       | 48  |
| Qy | 27  | GlyVala-----ProGlylleLeuTyrValargThrAspIleGlyGlyMetTyrArg-Tr | 44  |
| Db | 49  | GGCGGCTGGGATCGGGTGTGGGAATGTGGGCAACACCATCTCGGGGTATCTACAACGTG  | 108 |
| Qy | 44  | pAspAlaAlaAsnGlyArgTrpIleProLeuLeuAspTrpValGlyTrpAsnAsnTrpGl | 64  |
| Db | 109 | GGCAGCTGGACCTCTCGACCGCGCGCTTAACCTCG-----GGTGGCAA-----        | 154 |
| Qy | 64  | yTyrAsnGlyValValSerIleAlaAspProIle-----AsnThrAs              | 79  |
| Db | 155 | -----ATATCGGAACCAATATTTCGCGGCTGTCGCGACGGC                    | 192 |
| Qy | 79  | nLysVal-----TrpAlaAlaValGlyMetTyrThrAsnSerTrpAspPr           | 94  |
| Db | 193 | GGCGGTACTCGCGCTATTAACTTGGGCTTGGCCCAACCCAGCAACCTCAACGTGGGCTTC | 252 |
| Qy | 94  | oAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThrProLe | 114 |
| Db | 253 | GCAAGTCTCGGCGCTTAACTTCG-----                                 | 277 |
| Qy | 114 | uProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeuAlaValas | 134 |
| Db | 278 | -----GCGGCGCC  | 285 |
| Qy | 134 | pProAsnAsnAspAsnIleLeuTyrPheGlyAlaPro---SerGlyLysGlyLeuTrpAr | 153 |
| Db | 286 | ACCATCGGCCACCAACACGTGCGCATCGGGAACACCGGAATCTTCGATGTCGCGCTGGCG | 345 |
| Qy | 153 | gSer-----ThrAspSerGlyAlaThrTrpSerGlnMe                       | 164 |
| Db | 346 | AACCTGGGAGCTACAACTCGGCTTCGGAACCTTGGCGACGACCAACCTGGGCTTCGGC   | 405 |
| Qy | 164 | tThrAsnPheProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGl | 184 |
| Db | 406 | AACCTCG-----GCAGCTACACATCGGCTTCGGCAACGTGCGCAACG-----         | 448 |
| Qy | 184 | nSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer-----          | 199 |
| Db | 449 | -----ACAATCTGGGTTTCGCTAACGCGGCGCGGCGCAACATCGGCTTT            | 492 |

|    |      |  |      |
|----|------|--|------|
| Qy | 200  | -----SerSerLeuGlyGlnAlaSerLysThrIlePheValGl                  | 212  |
| Db | 493  | CGGAACACCGGCAGCAACAATGTCTGGTTCGGAAACCGGCAGCAACAATGTCTGGCAATC | 552  |
| Qy | 212  | yValAlaAspProAsnAsnProValPheTrpSerArgAspGlyGlyAla-----       | 228  |
| Db | 553  | GGGCTCACGGGCAACGGAC-----AGATCGGGTTCGGCAGCTTCAACTCG           | 597  |
| Qy | 229  | -----ThrTrpGlnAlaValPr                                       | 234  |
| Db | 598  | GGCAGCGGAACATCGGCTGTTCACTCGGCAGCAACAACATCGGATTCTTCAATTC      | 657  |
| Qy | 234  | oGlyAlaProThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLe | 254  |
| Db | 658  | GGCAGCGCAACTTCGGC-ATCGCAAACTCGGCAGCTTCAAC-----               | 699  |
| Qy | 254  | uTyrIleAlaThrSerAsnThrGlyGlyProTyrAspGly-----SerSerGlyAspVa  | 272  |
| Db | 700  | -----ACTGGCATCGGAACACCGGCACACCAATACCGGCTTATTCACCTCGGCGAC     | 753  |
| Qy | 272  | lTrpLysPheSerValThrSerGlyThrTrp-ThrArgIle-----SerProValProS  | 290  |
| Db | 754  | -----GTCAACACGGGCGCTTCAACCGGCGAGCTTCAACACCGGTAGCTT           | 800  |
| Qy | 290  | erThrAspThrAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisP | 310  |
| Db | 801  | CAACACCGGACGCTTCAACACCGGTGGCT--TCAATCGGCGCAATACCAACACCGGCTA  | 857  |
| Qy | 310  | roAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIleIle-----P  | 328  |
| Db | 858  | CCTCAACATTCGCAACTACAACACCGGCATCGCAACACCGGCGAGCTTCAACACCGGGC  | 917  |
| Qy | 328  | heArgSerThrAspGlyGlyAlaThrTrp-----ThrArgIleTrpA              | 342  |
| Db | 918  | TTTCATCACCGGAACCTACAGCAACGGGTGTGTTTAAGCGCGCATACCAACGGGCTTG-  | 976  |
| Qy | 342  | spTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProT | 362  |
| Db | 977  | -----TCGGCTCAACCT  | 989  |
| Qy | 362  | rpLeuThrPheGlyValGlnPro-----A                                | 370  |
| Db | 990  | GGTATCGATATGCTCTCTCCCAATAAGCTCGGCGTGAATATTCCCATCGATATCCGAT   | 1049 |
| Qy | 370  | snProProValProSerProLysLeuGlyTrp-----                        | 380  |
| Db | 1050 | CACGCTCGGCGCGCAACATCACCTTATGGCGCTCAGATTCCGCCACCGCGGATAT      | 1109 |
| Qy | 381  | -----MetAspGluAlaMetAlaIleAspProPheAsnSerA                   | 393  |
| Db | 1110 | CGTCCTTTCGTCATAGCGGCGCAGCGCCACTTTGGCGCCCATACATTCCGAACAT      | 1169 |
| Qy | 393  | spArgMetLeuTyr--GlyThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTr | 412  |
| Db | 1170 | CACGTTGTTCGCCCCCAGCAGCAGTCGCATAGGAGGCGCATATACCGCATCACCAT     | 1229 |
| Qy | 412  | pAspSerGlyGlnIleHisIle-----AlaProMetValLysGlyLeuGl           | 428  |
| Db | 1230 | AACCTGGGCTGGCGCATAGGATCCCGCTCATCAGTATCCGCGGCGCGCAGGTTTCGG    | 1289 |
| Qy | 428  | uGluThrAlaValAsn-----  | 433  |
| Db | 1290 | AACTCGACCAACCAACCGCTCGTCAGGTTTCTTCAATACCGCGCGCGCGCTTCGGG     | 1349 |
| Qy | 434  | -----AspLeuIleSerProProSerGl                                 | 441  |
| Db | 1350 | CTTCGGCAACTTCGGCGGCGCAATTCGGGCTTTTGGAACTTCGGCTTCGGACCTCGG    | 1409 |
| Qy | 441  | yAlaProLeuIleSerAlaLeuGlyAspLeuGly-----                      | 452  |
| Db | 1410 | GGCGTGGGCTCTCTCAACGTCGCGCGCTGGATCAGGTCTGGGGAACGTGGGCGCACAC   | 1469 |

```
Qy 453 -----GlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerPr 470
Db 1470 CGTCTGGGGTCTTACACACAGCAGCTCGGACCTCGGACCGCGGCTTCAATTACAG 1529
Qy 470 oValPheThrThrGlyThrSerVal----- 478
Db 1530 CTTGGCCCAACATCAGCACCATGATCGCGCGCTGCTGCGGACAGCAGCGGCACCATGGT 1589
Qy 479 -----AspTyrAlaGluLeuAsnProSerIleIleValArgAlaGl 492
Db 1590 CCTCAACCTGGGCTTGGCAACACCGCACCTCAACGTCGCAAT----- 1635
Qy 492 ySerPheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGl 512
Db 1636 -----GCAACCTCGGCGATACAAACATCGGCTTTGCAAACTCTCGGCAG 1679
Qy 512 yLysAsnTrp-----PheGlnGlySerGluProGlyGlyValThrTh----- 526
Db 1680 CGCCAACTTCGGCAGCGCAATATCGGTGGCAACACATCGGCGCGCAACACCGGAAT 1739
Qy 527 -----GlyGly-ThrValAlaAlaSerAlaAspGlySerA 538
Db 1740 ATTGACATCGTGTGGCAATCTGGCGAGTTACAACTCGGCTTCGGAACTTCGGCGA 1799
Qy 538 rgPheValTrpAlaProGlyAspProGlyGlnProValValThrAlaValGlyPheGlyA 558
Db 1800 TGCAACCTCGGCTTC-GGAACCTCGGCAGC-----TACAACCTCGGCTTCGGAA 1849
Qy 558 snSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValA 578
Db 1850 ACTTGGGCAACGACAACTCGGCTTC-----GCCAACACCGGC----- 1887
Qy 578 snProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerThrAspGlyGlyV 598
Db 1888 -----AGCAACATATCGGCTTCGGAACTCGGCAGC----- 1920
Qy 598 alThrPheGlnProValAlaGlyLeuProSerSerGlyAlaValGlyValMetPheH 618
Db 1921 -----AACAATATCGGCTTCAGCGGCGACGCGCAGATCGG-----TTTCG 1966
Qy 618 isAlaValProGlyLysGluGlyAspLeuTrpLeu-----AlaAlaSerSerG- 634
Db 1967 GCTCCTCGAATCTTGGCAGCGGAACATCGGCTTGTTCAACTCGGCGCAGCGAAACATCG 2026
Qy 634 lyLeuTyrHisSerThrAsnGly----- 641
Db 2027 GCTTTTCACTCGGCGAACGGAACAGTTGGCATCGGCAACACCGGCACCGCAACTTCG 2086
Qy 642 -----GlySerSerTrpSerAlaIleThrGlyVal-----SerSerAlaValAsnValG 658
Db 2087 GGCTTGGAAACACCGGACGACCAACACCGGCTTCTTCACTCGGCGCAGCTCATATACG 2146
Qy 658 lyPheGlyLysSerAla-----ProGlySerSerTrpProA 670
Db 2147 GTATCGGCAACACCGGACGCTTCAACACCGGCGAGCTTCAATCGGCGGATTCACACACG 2206
Qy 670 laValPhe-----ValValGlyThrIleGlyVal-----ThrG 682
Db 2207 GGGATTTCAACCCAGCGAGCTTACACACGGGATCGGAAACACCGCGGATGTTCACACG 2266
Qy 682 lyAlaTyrArgSerAspAspCys-----GlyThrThrTrpValLeuIleAsnAspAspG 700
Db 2267 GGGCTTCATCTCGGCGAGCTACAGCAACCGGTTCTTGTGG----- 2307
Qy 700 lnHisGlnTyrGlyAsnTrpGlyGlnAlaIleThrGlyAspHisAlaAsnLeu 717
Db 2308 -----AGTGGAAATTTAT---CAGGCGCTCATTTGGCTTTCACGCGCGCTA 2349
```

## RESULT 9

US-10-522-037-2/c

; Sequence 2, Application US/10522037

; Publication No. US20050282166A1

; GENERAL INFORMATION:

```
; APPLICANT: LIBRAGEN
; TITLE OF INVENTION: Method for the expression of unknown environmental DNA into adapted
; FILE REFERENCE: B0149WO
; CURRENT APPLICATION NUMBER: US/10/522,037
; CURRENT FILING DATE: 2005-01-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 37507
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: DNA sequence of clone FS3-135.
US-10-522-037-2
```

## Alignment Scores:

|                        |        |               |       |
|------------------------|--------|---------------|-------|
| Pred. No.:             | 0.107  | Length:       | 37507 |
| Score:                 | 161.00 | Matches:      | 181   |
| Percent Similarity:    | 31.0%  | Conservative: | 67    |
| Best Local Similarity: | 22.6%  | Mismatches:   | 277   |
| Query Match:           | 4.0%   | Indels:       | 276   |
| DB:                    | 8      | Gaps:         | 44    |

US-09-917-376-3 (1-740) x US-10-522-037-2 (1-37507)

|    |       |  |       |
|----|-------|--|-------|
| Qy | 36    | ThrAspIleGlyGlyMetTyrArgTrpAspAlaAlaAsnGlyArgTrpIlePro-----  | 53    |
| Db | 26051 | ACCGGACCTCGGCTCGCTCAGCTGG-----GCCAACGGCAGACCCAGACC           | 25998 |
| Qy | 54    | -----LeuLeuAspTrpValGlyTrpAsnAsnTrpGlyTyrAsnGlyVal           | 68    |
| Db | 25997 | ATCACGGTCCCGTGGCGGACGACGCTATCGGTGAAGGCAACGAGACGTTCAACGTCGCG  | 25938 |
| Qy | 69    | ValSerIleAlaAlaAspProIleAsnThrAsnLysValTrpAlaAlaValGlyMetTyr | 88    |
| Db | 25937 | CTCTCAATGCGACG-----GGTTCCACGGGTGCTCTTC                       | 25905 |
| Qy | 89    | ThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThr | 108   |
| Db | 25904 | ATCGGCGCGCGCGGACCGCGCATCATCTCTCG-----                        | 25866 |
| Qy | 109   | TrpGlnIleThrProLeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGly | 128   |
| Db | 25865 | -----CCGTGGCCACGGGTGGCAGCTGCTGCGGC-----                      | 25833 |
| Qy | 129   | GluArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSerGly | 148   |
| Db | 25832 | -----TTACGACCTCGGGCGGC                                       | 25815 |
| Qy | 149   | LysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhePro | 168   |
| Db | 25814 | GCCAAACGATGG---ACGGTCGGCTCCGACCGCTTCGAGCGCGCAGCAGCTTCGT      | 25758 |
| Qy | 169   | AspValGlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSerAspIleGln    | 188   |
| Db | 25757 | TCGGCGCAGGTGTGACGAGCGACCGACCACTCCGACGCGTGAACCTCGACCTCTCG     | 25698 |
| Qy | 189   | -----GlyValValTrp-----ValAlaPheAspLysSerSerSerLeuGly         | 203   |
| Db | 25697 | TTTCGGCGGACGCTGCTCGCGGCAACGCTCGCTTCGCCTATCGCGTTCGAGCTTCCCC   | 25638 |
| Qy | 204   | GlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnProValPheTrpSer | 223   |
| Db | 25637 | AACGCGCGCAAGTTCGAGTTCTCTG-----ATCGACGCGCAGGTGTTTCAGC         | 25590 |
| Qy | 224   | ArgAspGlyGlyAlaThr-----TrpGlnAlaValProGlyAlaProThrGlyPhe     | 240   |
| Db | 25589 | -----GACCGCGGACACACCGCTCGGCTCGAAGCTGTCGTCGCGGTGACCGCGG-      | 25534 |
| Qy | 241   | IleProHisLysGlyVal-----PheAspProValAsnHisValLeuTyrIleAla-Th  | 258   |
| Db | 25533 | CGCGCACACGCTGGTGTGGCGTTTCTCGAACTTCCTCTCCACGCTTGCAACTCGCGAC   | 25474 |

```
Qy 258 rSerAsnThrGlyGlyProTyrAspGlySerSerGlyAsp---ValTrp----- 273
Db 25473 CTGAAACCCGCGCGCCGATGCGCAACTGCGCGACCGCGCTGGATCGACGGGCT 25414
Qy 274 -LysPheSerValThrSerGlyThrThrTrpThrArgIleSerProValProSerThrAspTh 293
Db 25413 GGCACCTCCCGGTTCGGGTGGGTGTCACCGCGCCCAACCAAGCCGTCTCTCG---GACCT 25357
Qy 293 rAlaAsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIle 313
Db 25356 GAGCGCGAC-----GGCAGGAGCGACCTCTGCTG---CAGAACAGCGACGGCGCAT 25306
Qy 313 eMetValAlaThrGln----- 318
Db 25305 CGCGGTGTGGACATGAACGCGACCGCATCACCAGCTGCGGCCAACCTGATCGCGCGCG 25246
Qy 319 -----IleSerTrpTrpProasp-----ThrIleIle 327
Db 25245 CGCGGGTGGACGTGAGCGAGCGCGCGACCTCAATGGCGACCTGCGCGCGACATCCT 25186
Qy 327 ePheArgSerThrAspGlyGlyAlaThrThrTrpThrArgIleTrpAspTrpThrSerTyrPr 347
Db 25185 GTTCGCGCACCGATGGC----- 25167
Qy 347 oAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGlyVa 367
Db 25166 ---CGCGCTACGCGTACATCATGAAACGGGACTACG-----GTGACGGCGCGCAA 25120
Qy 367 lGlnProAsnProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIle 387
Db 25119 GGAGCTGCTGCTCC-----GCCGGCTGCGGTGATCTGTAGCCACACGCGC-- 25074
Qy 387 eAspProPheAsnSerAsp----- 393
Db 25073 ---GACCTGAACGCGCAGCGAAGCGGACATCATCTCGGCCACACCGAGCGACGCGC 25018
Qy 394 -----ArgMetLeuTyr-----G1 398
Db 25017 GCACATCTGTTGATGACGCGCAGCATCATCGGCGCGCCAGCTCTTCGGGGCGG 24958
Qy 398 yThrGlyAlaThrLeuTyrAlaThrAsnAspLeuThr----- 410
Db 24957 GACCGGTGACGCTCACCAGCGCGCGACATGACGCGCGACGCGACCTGGT 24898
Qy 411 -----LysTrpAspSerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuG1 428
Db 24897 CTTTCATGCACGCGCGCGCGCTACATCTATCATC-----ATGACGCGCACCC 24847
Qy 428 uGluThrAlaValAsnAspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLe 448
Db 24846 GGTGACGCGCGAGCGGCTTCTCTCGCGCGCGCGGTGG---ACGTAGCCACCT 24790
Qy 448 uGlyAspLeuGlyGly-----PheThrHisAlaAspValThrAl 461
Db 24789 GGGCGACCTCAACGCGCGCAGCGCGGACATGATCTTCGCGCACACCGACGCGCGCGC 24730
Qy 461 aValProSerThrIlePheThrSerProValPheThrThrGlyThrSerValAspTyrAl 481
Db 24729 G-----CACCTGTCTGTATGAACGCGCACCATCTTCGCGCGCAGC-----GC 24688
Qy 481 aGluLeuAsnProSerIleIleValArgAlaGlySer-----Ph 494
Db 24687 CAGCTCTCTCCG-----GCCGCGCGGGGTGGTGGTGAACCTGGTGGG 24643
Qy 494 eAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLysAs 514
Db 24642 CGACCTGAACGCGCGACGGAAGGACGACCTGTCTTCGCGCACACCGACGCGCGGTGCA 24583
Qy 514 nTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSerAl 534
Db 24582 C-----ATCCGCGCTGAT 24571
```

```
Qy 534 aAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAlaVa 554
Db 24570 GGATGCTGTTTCATCATCATCCGCG---GGCGAT-----GCGCT 24535
Qy 554 lClyPheGlyAsnSerTrpAlaAsaSerGlnClyValProAlaAsnAlaGlnIleArgSe 574
Db 24534 GCCGCGGGCACGGGTTGGAACGTACGACGCTGTGGACCTGAACCGGACGACGCGCAA 24475
Qy 574 rAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSerTh 594
Db 24474 GGACCTGTGTC-----TTCGCGCACAC 24454
Qy 594 rAspGlyGlyValThrPheGlnProValAlaAlaGlyLeuProSerSerGlyAlaValG1 614
Db 24453 CGACGCGACGCTCACGGTCCG---CTGATGAACGGGCTCGCCACCATCGGACGCGCAA 24397
Qy 614 yValMetPheHisAlaValProGlyLysGluGlyAspLeuTrpLeuAlaSerSerG1 634
Db 24396 CTTGATC----- 24390
Qy 634 yLeuTyrHisSerThrAsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAl 654
Db 24389 -----GCCCGCGCGGTGTCGTC-----GC 24367
Qy 654 aValAsnValGlyPheGlyLysSerAlaProGlySerSerTyrPro-----AlaVa 671
Db 24366 ACCCGCGCAACCTGAGCTTGGCGGTTCCCTCCACGGCGAGGCCCGCGCTCGCGCT 24307
Qy 671 lPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSerAspCysGlyTh 691
Db 24306 TTTT-----CATGGCGCGCGGCTTATAGACTGCGCGGAGACGTCC 24265
Qy 691 rThrTrpValLeuIleAsnAspAspGlnHisGlnTyrGlyAsnTrpGly-----G1 708
Db 24264 CTGGTGG-----TCCCATGATCTCGACGACACGCGCTGTTGGCGGTTCTCTCG 24214
Qy 708 nAlaIleThrGlyAspHisAlaAsnLeuArgValTyrIleGlyThrAsnGlyArg 727
Db 24213 CGGTGCTGCTGGCGGCGACGCTGGTCTCGGTGCTCGCTACGCGGTGCTGCAGCGCGC 24156
RESULT 10
US-10-793-626-3881
; Sequence 3881, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3881
; LENGTH: 3155
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3881
```

Alignment Scores:  
Pred. No.: 0.019 Length: 3155  
Score: 157.00 Matches: 168  
Percent Similarity: 34.5% Conservative: 80  
Best Local Similarity: 23.4% Mismatches: 299  
Query Match: 3.9% Indels: 172  
DB: 8 Gaps: 41

US-09-917-376-3 (1-740) x US-10-793-626-3881 (1-3155)



QY 27 GlyAlaPro-----GlyIleLeuTyrValArgThrAspIleGlyGlyMetTyrArgTrp 44  
DB 23 GGATCTCATCACTGGACCAATATTGTGCATCATCCACTGGGGTGG-----ACTACT 79  
QY 45 AspAlaAlaAsnGlyArgTrpIleProLeuLeu-----AspTyrValGlyTrp 60  
DB 80 TCGCCTGTATCAGGATTTTAACTCTCGTTTACCTGGAACTGCTCTTGGCTACCTTTC 139  
QY 61 AsnAsnTrpGlyTyrAsnGlyValVal-----SerIleAlaAlaAspPro----- 75  
DB 140 GGTGCAATTGGATCAAAATTCATCTTATGGCTGGCTGTGATTTCTTCGCCACCAATAAGA 199  
QY 76 ---IleAsnThrAsnLys-----ValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 92  
DB 200 ACGAFTTTCATCTACTGGTGTGTTTGTGTTTCTTGTGTTTCACTTGGCTACCTTTC 259  
QY 93 AspProAsnAspGlyAlaIleLeuArgSerAspGlnGlyAlaThrTrpGlnIleThr 112  
DB 260 TCCCTCTGTAAAGGTTCTTGTAGTT-----GTTGGTGTGTAAATTGTTTGTGT 307  
QY 113 Pro---LeuProPheLysLeuGlyGlyAsnMetProGlyArgGlyMetGlyGluArgLeu 131  
DB 308 CCGTGTTCACCTTTTGAACGACTTCTCTGTACCTGGCGCTAAGTTTGGATCAAAATTCG 367  
QY 132 AlaValAspProAsnAsn---AspAsnIleLeuTyrPheGlyAlaProSer---GlyLys 149  
DB 368 CGTTTTTTATCAACCGGAATTTCTTCGTTACGCTGATCGGATCTCCATCAACTGGACCA 427  
QY 150 GlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhe----- 167  
DB 428 TATTTTGTGCATCATCCACTGGTGGTGTGCTACTTCCGCTGTATCAGGTTTAACT 487  
QY 168 ---ProAspValGlyThrTyrIleAlaAsnProThrAspThrThrGlyTyrGln 184  
DB 488 CCTGTTTACTCGGAACGCTCTCTGGCTA-----CCTTTCGGTGCATTTGGATCAAAAT 541  
QY 185 SerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer----- 199  
DB 542 TCATCCCTTA-----TGGCTCGCTGTGATTTCTTCGCCACCAATAATGAACGATT 589  
QY 200 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---Asn 218  
DB 590 TCATCCACTGCT---TGTGTTTGTACTTTTCTGTTGTTTCACTTCCGCCCACTTTTTC 646  
QY 219 ProValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro--- 237  
DB 647 CCGTTAATAGGGTCTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 706  
QY 238 ---ThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIle 256  
DB 707 TGAACGACTTCTCTGTACTCGCGCTAAGTTTGGATCAAAATTCGGTTTTTATCAAAAC 766  
QY 257 AlaThrSerAsnThr-----GlyGlyProTyrAspGlySerSerGlyAspValTrp 273  
DB 767 GGAATTTCTCCGTGTACGTAATCGGATCTCCA-----TCAACTGGACCAATATTTT 817  
QY 274 LysPheSerValThrSerGlyThrTrpThrArgIleSerProVal----- 288  
DB 818 GTCACATCATCCACTGGTGGAGTAACACT-----TCGCGCTGTATCAGGATTTTAACT 871  
QY 289 ProSerThrAspThrAlaAsnAspTyr-----PheGlyTyrSerGlyLeuThrIleAsp 306  
DB 872 CCGTGTTTACTGGACCTCTCTGGCTACTTTTCGGTGCATTTGGA----- 919  
QY 307 ArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAspThrIle 326  
DB 920 -----TCAAAATTCATCTTATGSCCTGGCTTGTATT 949  
QY 327 IlePheArgSerThrAspGlyGlyAlaThrTrpThrArgIleTrpAspTrpThrSerTyr 346  
DB 950 -----TCTTCGCCA 958  
QY 347 ProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluProTrpLeuThrPheGly 366

DB 959 CCA-----TATCTGGATTTCACTACTGGTGTGTTTGTACTTTTCT 1003  
QY 367 ValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAla 386  
DB 1004 GTTGTGTTCACCTTCGCCAACTTTTCCCTGTTAATGGGTCTTCTAGTTGTGTGTGTA 1063  
QY 387 Ile-----AspProPheAsnSerAspArgMetLeuTyrGlyThrGlyAla 401  
DB 1064 ATTGTTTTTGTTCCTGGTTTCACTTCTGTTTAAACGCGCTCTTCTACCTGTTTTAAATCA 1123  
QY 402 ThrLeu-----TyrAlaThrAsnAspLeuThrLysTrpAsp----- 413  
DB 1124 GGAATTGAATTCAGTTTCTTGTGGAATGGAATTTCTTCGTTGACGTGATCGGATCTCCA 1183  
QY 414 SerGlyGlyGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsn 433  
DB 1184 TCACCTGGACCATATTTGTGCATCATCCACTGGTGGTGTGACTACTTCCGCTGTATCA 1243  
QY 434 AspLeuIleSerProProSerGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGly 453  
DB 1244 GGGTTTTTAACTCCT-----GGTTTACCTGGAACTCTCTTGGCTACCTTTCCGTTGCA 1297  
QY 454 PheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProValPheThr 473  
DB 1298 TTT---GGATCAAAATTCATCTTATGGCTGCTGATTTCTTCCGCCACCATAT----- 1348  
QY 474 ThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAlaGlySer 493  
DB 1349 -----TCTGTGATTTCACTACTGGTTGT 1372  
QY 494 PheAspProSerSerGlnProAsnAspArgHisValAlaPheSerThrAspGlyGlyLys 513  
DB 1373 TTT-----GTTATTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1402  
QY 514 AsnTrpPheGlnGlySerGluProGlyGlyValThrThrGlyGlyThrValAlaAlaSer 533  
DB 1403 CCAACTTTT---TCCCTGTTAATGGTCTCTTATGTTGTTGTTGTTGTTGTTGTTGTTT 1456  
QY 534 AlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProValValTyrAla 553  
DB 1457 GTTCTGTGTTCACTTTTGTGTTTAAACACGCTCTTCACTGGT-----TTTAAA 1504  
QY 554 ValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArg 573  
DB 1505 TCAGGATG---AATTCAGTTTCTTGTGCAATGGAATTTCTTCCGTTGACGTGATCGGA 1561  
QY 574 SerAspArgValAsnProLysThrPheTyrAlaLeuSerAsnGlyThrPheTyrArgSer 593  
DB 1562 TCTCCATCAACTGGACCA-----TATTTTGTACATCA 1594  
QY 594 ThrAspGlyGlyValThrPheGlnProValAlaAla-----GlyLeuProSer 609  
DB 1595 TCCACAGGTGGAGTAACACTTTCGCTGTATCAGGATTTTAAACCCCGCTTACCT--- 1651  
QY 610 SerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLeuTrp--- 628  
DB 1652 ---GGTTCGTTGTTGATCTACTTTCGCTGCAATTTGGATCAAAATTCATCTTATGCGCT 1708  
QY 629 -----LeuAlaAlaSerSerGlyLeuTyrHisSerThr 639  
DB 1709 GGCTTGATTTCTTCGCCACCAATAATGAACGATTTCACTCCACTGGTGTGTTGTACTTTT 1768  
QY 640 AsnGlyGlySerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGly--- 658  
DB 1769 TCTGTGTTGTTCACTTCGCCAACTTTTCCCTGTTAATGGGTCTTCTAGTTGTGTGTT 1828  
QY 659 -----PheGlyLysSerAlaProGlySerSerTyrProAlaValPheValVal 674  
DB 1829 GTAATTTGTTTTT-----GTTCTGTTTCACTTTTGAACGACGCTTCTCTGTA 1876  
RESULT 11  
US-10-793-626-3107/c

```

; Sequence 3107, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P0348005
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3107
; LENGTH: 2733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3107

Alignment Scores:
Pred. No.: 0.0215 Length: 2733
Score: 155.50 Matches: 164
Percent Similarity: 32.8% Conservative: 89
Best Local Similarity: 21.2% Mismatches: 287
Query Match: 3.9% Indels: 236
DB: 8 Gaps: 39

US-09-917-376-3 (1-740) x US-10-793-626-3107 (1-2733)

Qy 84 AlaValGlyMetThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSer 103
Db 2285 TCTGTGTTCCACCTTCGCCAACCTTTCTCTGTAAGGTTCTTAGTT----- 2235

Qy 104 AspGlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGlyGlyAsnMet 122
Db 2234 ---GTTGGTGTGAATGTTTGTTCCTGGTTCCACCTTTTGAACGACACTTCTCTGTA 2178

Qy 123 ProGlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn---AspAsnIleLeu 141
Db 2177 CTGGCGCTAAGTTGGATCAAAATCGCGTTTTTATCAACGGAATTTCTTCGGTTGAC 2118

Qy 142 TyrPheGlyAlaProSer---GlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThr 160
Db 2117 GTAATCGATCTCCATCACTGACGACCATATTTTGTACATCATCCACTGGTGGTGACT 2058

Qy 161 TrpSerGlnMetThrAsnPhe-----ProAspValGlyThrTyrlleAla 175
Db 2057 ACTTCGCTGTATCAGGATTTTAACTCCTGGTTTACCTGGAACGCTCCTCTGGCTA--- 2001

Qy 176 AsnProThrAspThrGlyTyrlleGlnSerAspIleGlnGlyValValTrpValAlaPhe 195
Db 2000 ---CCITTCGGTGGCTCGGATCAAAATTCAPCCTTA-----TGCCCTGGCTTG 1956

Qy 196 AspLysSerSer-----SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyVal 213
Db 1955 ATTTCITTCGCCACCATATTCGTGATTCATCTACTGTTGTTTGTATTTTCTGTT 1896

Qy 214 AlaAspProAsnAsnProValPheTrpSerArgAspGly-----GlyAlaThr 229
Db 1895 GGTTTCACCT---TGCCCAACTTTTCCCTCGTTAATGGTTCTTAGTTGTTGGTGGTTGA 1839

Qy 230 TrpGlnAlaValProGlyAlaPro-----ThrGlyPhe----- 240
Db 1838 ATTGTTTTTGTCTCGTTTCCACCTTCTGTTTAAACGCTCTTCACCTGGTTTAAATCA 1779

Qy 241 -----IleProHisLysGlyValPheAspProVal 250
Db 1778 GGATGAATTCACGTTTCTTGTGGAATGGAATTTCTTCGGTTGACGTAATCGGATCTCCA 1719

Qy 251 AsnHisValLeuTyrlleAlaThrSerAsnThrGlyGly----- 263

```

```

Db 1718 TCAACTGGACCATATTTTGTACATCATCCACTGGTGTGACTACTTTCGCTGTATCA 1559
Qy 264 -----ProTyrAsp-----Gly 267
Db 1658 GGAATTTTAACTCTCGTGTTTTACCTGGAAGCTCTCTTGGCTACCTTTCGGTGCATTGGA 1599
Qy 268 SerSerGlyAspValTrpLysPheSerValThrSerGly----- 280
Db 1598 TCAAATTCATCCTTATGGCTGGCTTGATTTCTCGCCACCATATAATGAACGATTTTCATCT 1539
Qy 281 -----ThrTrpThrArgIleSerProValProSerThrThrAspThrAlaAsn 295
Db 1538 ACTGGTGTGTTTGTGTACTTTTCTGTGTTCCACTTCGCCAACTTTTCCCTCGTTAAT 1479
Qy 296 AspTyrPhe----- 298
Db 1478 GGGTTCCTTAGTTGTGGTGTGTAATTTGTTTGTTCCTGGTTCACCTTTTGAACGACT 1419
Qy 299 -----GlyTyrSer 301
Db 1418 TTCTCTGTACCTGGCGCTAAGTTTGGATCAAAATTCGCGTTTTTATCAACGGAATTTCT 1359
Qy 302 GlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrp 321
Db 1358 TCCGTTGACGT-GATCGGATCTCCATCACTGACGACCATATTTTGTACATCATCCACTGG 1300
Qy 322 TrpProAspThrIle---IlePheArgSerThrAspGlyGlyAlaThrTrp-ThrArgIle 340
Db 1299 TGGTGTGACTACTTCGCTGTATCAGGTTTTTAACTCCTGGTTTACCTGGAACGCTCTC 1240
Qy 340 eTrp---AspTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAl 359
Db 1239 TTGGTACCTTTTGGTGCAATTTGGATCAAAATTCATCTTATGCGCTGGCTTGATTTCTTC 1180
Qy 359 aGluPro-----TrpLeuThrPheGlyValGlnProAs 370
Db 1179 GCCACCAATTAATGAACGATTTTATCCACCTGGTGTGTTTGTACTTCTCTGTGGTTCAC 1120
Qy 370 nProProValProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle----- 387
Db 1119 TTGCGCAACTTTTCCCTGTTAATGGTTCTTAGTTGTGGTGTGTGTAATTTGTTTGT 1060
Qy 388 -----AspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAlaThr 402
Db 1059 TCCTGGTTTCACCTTTTGAACGACTTTCTCTGTACCTGGCGCTAAGTTTGGATCAAAATTC 1000
Qy 402 xLeuTyrAlaThrAsnAspLeuThrLysTrpAsp-----SerGlyGlyGly 417
Db 999 GCGTTTTTTATCAACGGAATTTCTTCGCTTGACGTAATTCGGATCTCCATCAACTGGACC 940
Qy 417 nIleHisIleAlaProMetValLysGlyLeuGluGluThrAlaValAsnAspLeuIleSe 437
Db 939 ATATTTTGTACATCATCCACTGGTGGAGTAACCTTCGCTGTATCAGGATTTTAAAC 880
Qy 437 xPro-----ProSerGlyAla----- 442
Db 879 TCCTGGTTTACCTGGAACGCTCTCTTGGCTACCTTTCGGTGCATTTGGATCAAAATTCATC 820
Qy 443 -----ProLeuIleSerAlaLeuGlyAspLeuGlyGlyPh 454
Db 819 CTTATGGCTGGCTGATTTCTTCGCCACCATATTTCTGTGATTTTCATCTACTGTTGTTT 760
Qy 454 eThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal----- 471
Db 759 TGTACTTTTCTGTGTTGTTCA---CCTTCGCCAACTTTT---TCCCTGTGTTAATGGGTT 706
Qy 472 -PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleIleValArgAl 491
Db 705 CTTAGTTGTGGTGTGTAATTTGTTTGTTCCTGGTTACCTTTCTCTGTAAACGCGCTC 646
Qy 491 a-----GlySerPh 494
Db 645 TTTACCTGGTTTTAAATCAGGATTCAAATTCAGTTTCTTGTGCAATGGAATTTCTTCCT 586

```

QY 494 ensp-----ProSerSerGlnProAsnAspArgHisValAlaPheSerThrAs 510  
 DB TSAGCTGATCGGATCTCCATCACTGGACCA-  
 QY 510 pGlyGlyLysAsnTrpPheGlnGlySerGluProGlyValThrThrGlyGlyThrVa 530  
 DB -----TATTTTGTGCATCATCCACTGGTGTGTGACTACTTCGCCTGTA-- 510  
 QY 530 lAlaAlaSerAlaAspGlySerArgPheValTrpAlaProGlyAspProGlyGlnProVa 550  
 DB -----TCAGGGTITTTA--ACTCTCTGGTTTACCTGGAACG----- 477  
 QY 550 lValTrpAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGlyValProAlaAsnAl 570  
 DB -----TCTCTTGGCTACCTTTCCTGGTGCATTTGGATCAAAATTC 439  
 QY 570 aGlnIleArgSerAspArgValAsnProLysThrPheTrpAlaLeuSerAsn--GlyTh 589  
 DB ATCCTTATGGCCTGGTGTGATTTCTTCCGCCACCATATTTCTGTGATTTTCATCTACTGTTG 379  
 QY 589 rPheTrpArgSerThrAspGly-----GlyValThrPheGlnProValAlaAlaGlyLe 607  
 DB TTTTGTATTTTCTGTGTGTTTTCACCTTCCCAACTTTTCCCTGTATATGGTTCCTT 319  
 QY 607 uProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAspLe 627  
 DB -----GTTGTGTGTGTGTAATTTTCTTCTCTGTT-----TCACCTTT 277  
 QY 627 uTrpLeuAlaAlaSerSerGlyLeuTrpHisSerThrAsnGlyLysSerSerTrpSerAl 647  
 DB TTTGTTTAAACAGCTCTTTCACCTGTTGTTTAAATCA-----GGATTTGAATTCACGTTTCTT 223  
 QY 647 aIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySerSe 667  
 DB GTCCGATGGAAATTTCTTCGTTGAGTGATCGGATCTCCATCAACTGGACCA----- 171  
 QY 667 rTyProAlaValPheValValGlyThrIleGlyGlyValThr 681  
 DB -----TATTTTGTGCATCATCCACAGGTGGAGTAACT 138

RESULT 12

US-11-136-527-3719  
 ; Sequence 3719, Application US/11136527  
 ; Publication No. US20050287570A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wyeth  
 ; APPLICANT: Mounts, William M  
 ; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
 ; FILE OF INVENTION: 031896-041000 (AM101086)  
 ; CURRENT APPLICATION NUMBER: US/11/136,527  
 ; CURRENT FILING DATE: 2005-05-25  
 ; PRIOR APPLICATION NUMBER: US 60/574,294  
 ; FILING DATE: 2005-05-26  
 ; NUMBER OF SEQ ID NOS: 362830  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 3719  
 ; LENGTH: 2595  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-11-136-527-3719

Alignment Scores:  
 Pred. No.: 0.0224 Length: 2595  
 Score: 155.00 Matches: 163  
 Percent Similarity: 29.3% Conservative: 55  
 Best Local Similarity: 21.9% Mismatches: 276  
 Query Match: 3.8% Indels: 252  
 DB: 12 Gaps: 35

US-09-917-376-3 (1-740) x US-11-136-527-3719 (1-2595)

QY 12 AlalleGlyGlyGlyGlyPheValAspGlyIleValPheAsnGluGlyAla----- 28

DB 454 GCTGTGGTGGTATTGGTGGCATCGGTGGCTAGGAGTCTCAACAGGTGCTGTGGTCCCT 513  
 QY 29 -----  
 DB 514 CAACCTGGAGCTGGAGTGGAGCCGAGGAAAGCCTGGGAAAGTTCCTGTGTGC----- 567  
 QY 34 ValArgThrAspIleGlyGlyMetTrp--ArgTrpAspAlaAlaAsnGlyArg----- 50  
 DB -----GGTCTTCCAGGTGTATACCCAGGTGGAGTGTCTCCAGGAAACAGGAGTCCG 618  
 QY 51 -----TrpIleProLeuLeuAspTrp-----V 58  
 DB 619 TTCCTCTGTGTGGAGTGTCTCCTGGAGTTCCTCCACCTGGCACAGGAGTCAAGGCCAAGTT 678  
 QY 58 aLGIYTrpAsnAsnTrp--GlyTrpAsnGlyValValSerIleAla----- 72  
 DB 679 CCGGTGGAGGAGGTGGTCTTTTCTGGAATCCAGGGGTGGGCCCTTTTGGGGGTCTAG 738  
 QY 73 -----AlaAspProIleAsnThrAsnLysValTrpAlaAlaValGly 86  
 DB 739 CAGCCTGTGTCCACTGGTGTATCCCATCAAGCCCAAGCTGCCAGGTGGTATGGA 798  
 QY 87 Met---TyrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAspGln 105  
 DB 799 CTGCCCTATACCAATGG----- 816  
 QY 106 GlyAlaThrTrpGlnIleThrProLeuProPheLeuLeuGlyGly----- 120  
 DB 817 -----AAACTGTCCTATGAGTGGTGTGGTGGAGGGGCAAGCT 855  
 QY 121 AsnMetPro--GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsnAspAsn 139  
 DB 856 GGTATCCCAACAGGACAGGAGGTGGTCCAGGACAGTGGCAGCAGCTAAA---GCA 912  
 QY 140 IleLeuTrpPheGlyAlaProSerGlyLysGlyLeuTrpArgSerThrAspSerGlyAla 159  
 DB 913 GCGAAGTATGCAAGTGTCT---GGAGGAGGTGGAGTCTCCTGTGTGGAGGGGTGGC 969  
 QY 160 ThrTrpSerGlnMetThrAsnPheProAspValGlyThrTrpIleAlaAsnProThrAsp 179  
 DB 970 ATTCCTGTGTGTGGCGCAATCTCTGGATGGA----- 1005  
 QY 180 ThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSerSer 199  
 DB 1006 -----GGCATTCAGCGCTGGAACTCCCGCAGCAGCTGCTGTGTGCA 1047  
 QY 200 SerSerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsnAsnPro 219  
 DB 1048 AAGGTGCGCGCTAAGGCTGCTAAGTATGGAGCTGTGGAGGCTTAGTGCCTGTGGACCA 1107  
 QY 220 ValPheTrpSerArgAspGlyAlaThrTrpGlnAlaValProGlyAlaProThrGly 239  
 DB 1108 GGAGTTAGGTTCCAGGTGCTGGAATCCAGGTGTTGGATCCAGGT-----GTCGGT 1161  
 QY 240 PheIleProHisLysGlyValPheAspProValAsnHisValLeuTrpIleAlaThrSer 259  
 DB 1162 GGCATTCAGGTGTTGGTGGCATTCAGGTGTTGGTGGCATC-----CCA 1206  
 QY 260 AsnThrGlyGlyProTrpAspGlySerSerGlyAspValTrpLysPheSerValThrSer 279  
 DB 1207 GGTGTGGGGGCTGTATCGAGGTCCAGGATTTGTG----- 1245  
 QY 280 GlyThrTrpThrArgIleSerProValProSerThrAspThrAlaAsnAspTrpPheGly 299  
 DB 1246 GGTGACACAGGGGCTGTGTACAGCTGCAGCTGCCAAAGTCTGCCAAAGTGCACAG 1305  
 QY 300 TyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIle 319  
 DB 1306 TATGGA-----GCCAG-AGGTGG 1322  
 QY 320 SerTrpTrpProAspThrIleIlePheArgSerThrAspGlyAlaThrTrpThrArg 339



QY 241 -----IleProHisLysGlyValPheAspPro 249  
 DB 515 TCAGGATTGAATTCACGTTTCTTGCGAATGAATTTCTCGTTGACGTGATCGGATCT 574  
 QY 250 ValAsnHisValLeuTyrIleAlaThrSerAsnThrGlyGly  
 DB 575 CCATCACTGGACCAATATTTGTGCATCATCATCCACAGGTGGAGTAACACTTCGCGCTGTA 634  
 QY 264 -----ProTyrAsp----- 266  
 DB 635 TCAGGATTTTAAACCCCGGCTTACCTGTTGCGGTGTTTGACATACCTTTCGGTGCAATT 694  
 QY 267 GlySerSerGlyAspValTyrLysPheSerValThrSerGly-  
 DB 695 CGATCAAAATTCATCCTATGCGCTGCTGATTTCTTCGCCACCACCAATAAAGACGATTTC 754  
 QY 281 -----ThrTyrThrArgIleSerProValProSerThrAspThrAla 294  
 DB 755 TCCACTGGTGTGTTGTACTTTTCTGTTGCTTACCTTCGCCCAACTTTTCCCGCTGT 814  
 QY 295 AsnAspTyrPhe----- 298  
 DB 815 AATGGGTTCTAGTTGTTGGTGTGTAATGTTTGTTCCTGTTACCTTTTGAACG 874  
 QY 299 -----GlyTyr 300  
 DB 875 ACTTTCTCTGACCTGGCGTAAGTTTGGATCAAAATTCGCGTTTATCAAAACGAATT 934  
 QY 301 SerGlyLeuThrIleAspArgHisProAsnThrIleMetValAlaThrGlnIleSer 320  
 DB 935 TCTTCGTTGACGTAAT-CGGATCTCCATCACTGACGACCATATTTTGTACATCATCCAC 993  
 QY 321 TrpTyrProAspThrIle-IlePheArgSerThrAspGlyGlyAlaThrTyr-ThrAr 339  
 DB 994 TGGTGGAGTAATACTTCGCTGTATACGGTGTAACTCCTGTTTACCTGGACGTC 1053  
 QY 339 GileTyr---AspTyrThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSe 358  
 DB 1054 CTCCTGGTACTCTTCGGTGCATTTGGATCAAAATTCATCCATTATGCGCTGCTTCAATTC 1113  
 QY 358 rAlaGluPro-----TrpLeuThrPheGlyValGlnPr 369  
 DB 1114 TTCGCCACCAATAATGAACGATTTTCATCCACTGTTGTTTGTACTTTCTGTTGGTTC 1173  
 QY 369 oAsnProValProSerProLysLeuGlyTyrMetAspGluAlaMetAlaIle----- 387  
 DB 1174 ACCTTCGCCAACTTTTCCCGCTGTAATGGGTCTTACTGTTGTGTTGTAATTTGTTT 1233  
 QY 388 -----AspProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAl 401  
 DB 1234 TGTTCTCGTTACCTTTTGAACGACTTCTCTGACTGCGCTAAGTTTGGATCAAA 1293  
 QY 401 aThrLeuTyrAlaThrAsnAspLeuThrLysTyrAsp-----SerGlyGly 416  
 DB 1294 TTCGGGTTTTTATCAACAGGAATTTCTTCGCTTGACGTAATCGGATCTCCATCACTGG 1353  
 QY 416 yGlnIleHisIleAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuI 436  
 DB 1354 ACCATATTTTGTGCATCATCATCCACTGCTGGAGTAATACTTTCGCTGTATCAGGATTTT 1413  
 QY 436 eSerPro-----ProSerGlyAla----- 442  
 DB 1414 AACTCTCGTTTACCTGGAACGCTCTCTTGGCTACCTTTTCGGTGCAATTTGGATCAAAATTC 1473  
 QY 443 -----ProLeuIleSerAlaLeuGlyAspLeuGlyGly 453  
 DB 1474 ATCCTTATGGCTGCTGATTTCTTCGCCACCAATTTCTGTGATTTCTACTGTTG 1533  
 QY 453 yPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSerProVal----- 471  
 DB 1534 TTTTGTACTTTTCTGTGTTGGTTCA---CCTTCGCCAACTTTT---TCCCTGTTAATGG 1587

QY 472 ----PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSerIleValAr 490  
 DB 1588 GTTCTTAGTTGTTGGTGTGTAATTTGTTCTGCTGTTACCTTTCTGTTTAAACGCG 1647  
 QY 490 gAla-----GlySe 493  
 DB 1648 CTCCTTACCTGGTTTTAAATCAGGATTGAATTCACGTTTCTTGTGCAATGAATTTCTTC 1707  
 QY 493 rPheAsp-----ProSerSerGlnProAsnAspArgHisAlaValAlaPheSerTh 509  
 DB 1708 CGTTGACGTGCGGATCTCCATCACTGACCA----- 1741  
 QY 509 rAspGlyGlyLysAsnThrPheGlnGlySerGluProGlyGlyValThrThrGlyGlyTh 529  
 DB 1742 -----TATTTTGTACATCATCCACTGCTGTTGTTGACTACTTCGCTGT 1785  
 QY 529 rValAlaAlaSerAlaAspGlySerArgPheValTyrAlaProGlyAspProGlyGlnPr 549  
 DB 1786 A-----TCAGGGTTTTTA---ACTCCTGGTTTACCTCGAAGC-- 1819  
 QY 549 oValValTyrAlaValGlyPheGlyAsnSerTyrAlaAlaSerGlnGlyValProAlaAs 569  
 DB 1820 -----TCCTCTTGGCTACCTTTTCGGTGCATTTTCGATCAAA 1854  
 QY 569 nAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeuSerAsn--G1 588  
 DB 1855 TTCACTCTTATGCGCTGCTGATTTCTTCGCCACCATATTTCTGATTTTCACTACTGG 1914  
 QY 588 yThrPheTyrArgSerThrAspGly-----GlyValThrPheGlnProValAlaAlaG1 606  
 DB 1915 TTGTTTGTATTTTCTGTTGTTTCACTTCGCCAACTTTTTCCTGTTTAAATGGGTT 1974  
 QY 606 yLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGlyLysGluGlyAs 626  
 DB 1975 CTTA-----GTTGTTGGTGTGTAATTTGTTTGTCTCTGTT-----TCACC 2016  
 QY 626 pLeuTyrLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGlySerSerTrpSe 646  
 DB 2017 TTTTGTGTTTAAACAGCTCTTCACTGTTTAAATCA-----GGATTGAATTCACGTTT 2070  
 QY 646 rAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSerAlaProGlySe 666  
 DB 2071 CTTGTGCAATGAATTTCTTCGCTTACGATCGGATCGGATCTCCATCACTCACTGACCA----- 2125  
 QY 666 rSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGlyAlaTyrArgSe 686  
 DB 2126 -----TATTTGTACATCATCATCCACAGGTGGAGTAATACTTCGCTGTATC 2172  
 QY 686 rAsp-----AspCysGlyThrThrTyrValLeuIleAs 697  
 DB 2173 AGGATTTTAAACCCCGGCTTACCTGTTGCTGTTGTTGACTACCTTTCGGTGCATTTG 2232  
 QY 697 nAspAspGlnHisGlnTyrGly 704  
 DB 2233 -CATCAAAATTCATCTTATGCG 2253  
 RESULT 14  
 US-10-793-626-3970/c  
 ; Sequence 3970, Application US/10793626  
 ; Publication No. US2005025478A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KIMMERLY, WILLIAM JOHN  
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
 ; FILE REFERENCE: PUS480US  
 ; CURRENT APPLICATION NUMBER: US/10/793,626  
 ; CURRENT FILING DATE: 2004-03-04  
 ; PRIOR APPLICATION NUMBER: 60/164,258  
 ; PRIOR FILING DATE: 1999-11-09  
 ; NUMBER OF SEQ ID NOS: 4472  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3970  
 ; LENGTH: 3748  
 ; TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: nucleic acid sequence

US-10-793-626-3970

Alignment Scores:

Pred. No.: 0.0549 Length: 3748  
Score: 152.00 Matches: 162  
Percent Similarity: 33.0% Conservative: 91  
Best Local Similarity: 21.1% Mismatches: 294  
Query Match: 3.8% Indels: 220  
DB: 8 Gaps: 38

US-09-917-376-3 (1-740) x US-10-793-626-3970 (1-3748)

Qy 85 ValGlyMetThrThrAsnSerTrpAspProAsnAspGlyAlaIleLeuArgSerSerAsp 104  
Db 2516 GTTGGTTCACTTCACCAACTTTTCCCGTGTAAATGGTTCTAGTT----- 2469  
Qy 105 GlnGlyAlaThrTrpGlnIleThrPro---LeuProPheLysLeuGlyGlyAsnMetPro 123  
Db 2468 GTTGGCGTGTGAATGTGTTTGTCTCGTGTACCTTTTGAACGACTTTCTCTGTACCT 2409  
Qy 124 GlyArgGlyMetGlyGluArgLeuAlaValAspProAsnAsn---AspAsnIleLeuTyr 142  
Db 2408 GGGCGTAAGTTTGGATCAAAATTCGCGTTTATCAACGGAATTTCTTCGTTGACGTA 2349  
Qy 143 PheGlyAlaProSer---GlyLysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrp 161  
Db 2348 ATCGGATCTCACTCACTGACCACTATTTGTTCATCATCTGCTGGTGGTGGTACTACT 2289  
Qy 162 SerGlnMetThrAsnPhe-----ProAspVal---GlyThrTyrIleAlaAsnProThr 178  
Db 2288 TCGCGTGTATCAGGATTTTAACTCTCGTGTACTGTTGCGTGTGTGTACTACCTTTC 2229  
Qy 179 AspThrThrGlyTyrGlnSerAspIleGlnGlyValValTrpValAlaPheAspLysSer 198  
Db 2228 GGTGCGTTCGGATCAAAATTCATCCTTA-----TGGCCTGTCTGATTTCTTCG 2181  
Qy 199 SerSer-----SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspPro 216  
Db 2180 CCACCAATAAGACGATCTCATCCACCTGGTGTGTTTGTATTTTCTGTGTGGTTCACCT 2121  
Qy 217 AsnAsnProValPheTrpSerArgAspGly-----GlyAlaThrTrpGlnAla 232  
Db 2120 ---TCGCCAACTTTTCTCTGTTAATGGGTTCTTAGTTGTTGGTGTGTTAATGTTT 2064  
Qy 233 ValProGlyAlaPro-----ThrGlyPheIleProHisLysGlyValPheAspProVal 250  
Db 2063 GTTCTCGTTCACCTTTTGAACGACTTCTCTGTACCTGCGCTAAGTTTGGATCAAA 2004  
Qy 251 AsnHisValLeu----- 254  
Db 2003 TCGCGTGTATTAACAAACGGAATTTCTTCGTTGACGTAATTCGGATCTCCACCACTGA 1944  
Qy 255 ---TyrIleAlaThrSerAsnThrGlyGly----- 263  
Db 1943 CCAATTTTGTACATCACTCACTCGTGGTGTGACTACTTCGCGCTGTATCAGGATTTTA 1884  
Qy 264 -----ProTyrAsp-----GlySerSerGly 270  
Db 1883 ACTCTCGTGTACTCGAAGCTCTCTTGGCTACTTTTCGGTGGTTCGGATCAAAATCA 1824  
Qy 271 AspValTrp-----LysPheSerValThrSerGlyThrTrpThr 283  
Db 1823 TCCTTAGGCGCTGGCTTGATTTCTTCGCCACCAATATTCGTGATTTCTATCTACTGTTGT 1764  
Qy 284 ArgIle-----SerProValProSerThrAspThrAlaAsnAspTyrPhe 298  
Db 1763 TTTGTTATTTTCTGTGTGTTCACTTCGCCCAACTTTTCCCTGTGTTAATGGTCTCTTA 1704  
Qy 298 ----- 298

Db 1703 GTTGTGGTGTGAATGTTTGTCTCTGTTCCACCTTTCACCTTCTGTGTTAACACGCTCTTCA 1644  
Qy 299 -----GlyTyrSerGlyLeuThr 304  
Db 1643 CTGGTTTTTAATCAGGATTGAATTCACGTTTCTTGTGCAATGGAATTTCTTCCGTTGAC 1584  
Qy 305 IleAspArgGlnHisProAsnThrIleMetValAlaThrGlnIleSerTrpTrpProAsp 324  
Db 1583 GTAAT-CGGATCTCCATCACTCGGACCAATAATTTTGTACATCATCCACTCGTGGTGTGAC 1525  
Qy 325 ThrIle---IlePheArgSerThrAspGlyAlaThrTrp-ThrArgIleTrp---As 342  
Db 1524 TACTTCGCGTGTATCAGGATTTTAACTCCTCGTGTATCTGGAACGCTCTCTTGGCTACC 1465  
Qy 342 pTrpThrSerTyrProAsnArgSerLeuArgTyrValLeuAspIleSerAlaGluPro-- 361  
Db 1464 TTTCCGTCATTTGGATCAAAATTCATCTTATGCGCTGCTGTGATTTCTTCCGACCAATA 1405  
Qy 362 -----TrpLeuThrPheGlyValGlnProAsnProProVa 373  
Db 1404 ATGAACGATTTTCATCTACTGTTGTTTGTGTACTTTTCTGTGTGTTACCTTCGCGAAC 1345  
Qy 373 lProSerProLysLeuGlyTrpMetAspGluAlaMetAlaIle-----As 388  
Db 1344 TTTTCCCTCTGTTAATGGGTTCTTAGTTGTTGGTGTGTTGTAATGTTTGTCTCTGTTTC 1285  
Qy 388 pProPheAsnSerAspArgMetLeuTyrGlyThr-----GlyAlaThrLeuTyrAl 405  
Db 1284 ACCTTTGAACGACTTCTCTGCTACCTGGCGCTGAAGTTGGATCAAAATTCGCGTTTTT 1225  
Qy 405 aThrAsnAspLeuThrLysTrpAsp-----SerGlyGlyGlnIleHisI1 420  
Db 1224 ATCAACGGAATTTCTTCGTTGACGTGATCGGATCTCCATCACTCGGACCAATAATTGT 1165  
Qy 420 eAlaProMetValLysGlyLeuGluThrAlaValAsnAspLeuIleSerProSe 440  
Db 1164 CACATCATCCACTGGTGGTGTACTTCTGCTGTATCAGGTTTAACTCTCT----- 1110  
Qy 440 xGlyAlaProLeuIleSerAlaLeuGlyAspLeuGlyGlyPheThrHisAlaAspValTh 460  
Db 1109 -GGTTTACCTGGAACGCTCTCTGCTGCTACCTTTCGGTGCATTT---GGATCAAAATCATC 1054  
Qy 460 xAlaValProSerThrIlePheThrSerPro-----ValPheThrThrGlyThrSe 477  
Db 1053 CTATGCGCTGGCTGTGATTTCTTCGCCACCAATAATGAACGATTTTCATCCACTGGTGT 994  
Qy 477 xValAspTyrAlaGluLeuAsnProSer----- 486  
Db 993 TGTATCTTTTCTGTGTTTCACCTTCGCCAACTTTTTCCTGTTAATGGTCTCTTAGT 934  
Qy 487 -----IleIleValArgAlaGlySerPheAspProSerSerGlnProAsnAspAr 503  
Db 933 TGTGTTGTTGTAATGTT-----TTTGTCTCTGTTCA-----CCTTT 895  
Qy 503 gHisValAlaPheSerThrAspGlyGlyLys----- 513  
Db 894 TTGAACGACTTCTCTGTACCTCGCGCTAAGTTGGATCAAAATTCGCGTTTTTTTATCAA 835  
Qy 514 -----AsnTrpPheGlnGlySe 519  
Db 834 CGGAATTTCTTCGTTGACGTAATCGGATCTCCATCACTCGGACCAATAATTTGTACATC 775  
Qy 519 xGluProGlyGlyValThrThrGlyThrValAlaAlaSerAlaAspGlySerArgPh 539  
Db 774 ATCCACTGGTGGAGTAACACTACTTCGCGCTGTA-----TCAGGAT 736  
Qy 539 eValTrpAlaProGlyAspProGlyGlnProValValTyrAlaValGlyPheGlyAsnSe 559  
Db 735 TTTA---ACTCTGTTTACTCTGGAACG-----TCCTC 706  
Qy 559 rTipAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPr 579  
Db 579 rTipAlaAlaSerGlnGlyValProAlaAsnAlaGlnIleArgSerAspArgValAsnPr 579

Db 705 TTGGCTACCTTTTCGGTCATTTGGATCAAAATTCATCTTATGGCCTGGCTTGAATTTCTTC 646

Qy 579 olvysThrPheTyrAlaLeuSerAsn---GlyThrPheTyrArgSerThrAspGly----- 596

Db 645 GCCACCAATATCTGTCGATTTTCATCTACTCGTGTGTTTGTACTTTTCTGTTGGTTCACC 586

Qy 597 -GlyValThrPheGlnProValAlaAlaGlyLeuProSerGlyAlaValGlyValMe 616

Db 585 TTCGCCAACTTTTCCCTGTTAATGGTTCCTA-----GTTGTGGTGGTGTGT 538

Qy 616 tPheHisAlaValProGlyLysGlyAspLeuTrpLeuAlaAlaLaserSerGlyLeuTy 636

Db 537 AATGTTTGTCTCTGCTGCTACCT-----TTCGTTTAAACGGCTCTTTTACC 490

Qy 636 rHisSerThrAsnGlyLysSerSerTrpSerAlaIleThrGlyValSerSerAlaValas 656

Db 489 TGGTTTAAATCAGGATTTCAAGTTTCTTGTGCGAATGGAATTTCTTCGGTTGACGT 430

Qy 656 nValGlyPheGlyLysSerAlaProGlySerSerTyrProAlaValPheValGlyTh 676

Db 429 GATCGGATCTCCATCAACTGGACCA-----TATTTTGTACATCATC 388

Qy 676 rIleGlyGlyValThr 681

Db 387 CACTGGTGGTGTGACT 372

RESULT 15

US-10-793-626-4165/c

; Sequence 4165, Application US/10793626

; Publication No. US20050255478A1

; GENERAL INFORMATION:

; APPLICANT: KIMBERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: P03480US

; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4165

; LENGTH: 3600

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-4165

Alignment Scores:

| Pred. No.:             | Score: | Length:       | Matches: |
|------------------------|--------|---------------|----------|
| 0.0577                 | 151.50 | 3600          | 183      |
| Percent Similarity:    | 34.6%  | Conservative: | 91       |
| Best Local Similarity: | 23.1%  | Mismatches:   | 303      |
| Query Match:           | 3.8%   | Indels:       | 215      |
| DB:                    | 8      | Gaps:         | 46       |

US-09-917-376-3 (1-740) x US-10-793-626-4165 (1-3600)

Qy 27 GlyAlaPro-----GlyIleLeuTyrValArgThrAspIleGlyMetTyrArgTrp 44

Db 3600 GGATCTCCATCAACTGGACCATATTTTGTTCACATCATCCATCGTGGTGTG---ACTACT 3544

Qy 45 AspAlaAlaAsnGlyArgTrpIleProLeu-----AspTrpValGlyTrp 60

Db 3543 TCGCCTGTATCAGGGTTTTTAACCTCCTGTTTACCTGGAACGCTCTCTTGGCTACCTTTC 3484

Qy 61 AsnAsnTrpGlyTyrAsn-----GlyValValSerIleAlaAlaAspPro 75

Db 3483 GGTGCATTTGGATCAAAATTCATCCTTATGGCTGGCTTGAATTTCTTCGCCACCATATTTCT 3424

Qy 76 IleAsnThrAsnLys-----ValTrpAlaAlaValGlyMetTyrThrAsnSerTrp 92

Db 3423 GTGATTTTCATCTACTGTTGTTTGTATTTTCTGTTGGTTCACCTTCGCCAACTTTT 3364

Qy 93 AspProAsnAspGlyAlaIleLeuArgSerSerAspGlnGlyAlaThrTrpGlnIleThr 112

Db 3363 TCCCTGTTTAATGGGTTCTTAGTT-----GTTGGTGTGTAATTTGTTTGT 3316

Qy 113 Pro---LeuProPheLysLeuGlyGlyAsnMetProGly-----ArgGlyMetGlyGlu 129

Db 3315 CCTGGTTACCTTTTGTTTAAACACGCTCTCTACCTGGTTTTAAATCAGGATTAATTC 3256

Qy 130 ArgLeuAlaValAspProAsnAsnAspAsnIleLeuTyrPheGlyAlaProSer---Gly 148

Db 3255 CATTTC---TTCGCAATCGAATTTCTTCGGTTCAGCTGATCGATCTCCATCACTGGA 3199

Qy 149 LysGlyLeuTrpArgSerThrAspSerGlyAlaThrTrpSerGlnMetThrAsnPhe--- 167

Db 3198 CCATATTTTGTACATCATCCACAGGTGAGTAACCTACTTCGCTGTATCAGGATTTTA 3139

Qy 168 ---ProAspVal---GlyThrTyrIleAlaAsnProThrAspThrGlyTyrGlnSer 185

Db 3138 ACCCCCGGCTTACCTGGTTCGTTGTTGACTACCTTTCGGTGCATTTTCGATCAAAATTC 3079

Qy 186 AspIleGlnGlyValValTrpValAlaPheAspLysSerSer-----Ser 200

Db 3078 TCCCTTA-----TSGCCTGGCTTGAATTTCTTCGCCACCATAATGAACGATTTCA 3031

Qy 201 SerLeuGlyGlnAlaSerLysThrIlePheValGlyValAlaAspProAsn---AsnPro 219

Db 3030 TCCACTGGT---TGTTTGTACTTTTCTGTTGGTTCACCTTCGCCAACTTTTCCCT 2974

Qy 220 ValPheTrpSerArgAspGlyGlyAlaThrTrpGlnAlaValProGlyAlaPro----- 237

Db 2973 GTTAATGGTCTTCTAGTCTGTGTTGTTGTAATTTGTTTCTGCTGTTCCACCTTTTGA 2914

Qy 238 ThrGlyPheIleProHisLysGlyValPheAspProValAsnHisValLeuTyrIleAla 257

Db 2913 ACGACTTTCTCTGTACCTGGCGCTAAGTTTGGATCAAAATTCGCGTTTTTTATCAACCGGA 2854

Qy 258 ThrSerAsnThr-----GlyGlyProTyrAspGlySerSerGlyAspValTrpLys 274

Db 2853 ATTTCTTCGGTTGACGTAAATCGGATCTCCA-----TCAACTGGACCATATTTTGTGTC 2803

Qy 275 PheSerValThrSerGlyThrTrpTrpArgIleSerProValProSerThrAspThrAla 294

Db 2802 ACATCATCCACTCGTGGAGTAACCTACT-----TCGCTGTATCAGG--- 2761

Qy 295 AsnAspTyrPheGlyTyrSerGlyLeuThrIleAspArgGlnHisProAsnThrIleMet 314

Db 2760 -----TTTTAACTCCTCGTTTA-----CCTGGAACG----- 2734

Qy 315 ValAlaThrGlnIleSerTrpTrpProAspThrIleIlePheArgSerThrAspGlyGly 334

Db 2733 -----TCCTCTTGGCTACCT----- 2719

Qy 335 AlaThrTrpThrArgIleTrpAspTrpTrpSerTyrProAsnArgSerLeuArgTyrVal 354

Db 2718 -----TTCGGTGCATTTGGATCAAAATTCATCCTTATGCGCCT 2693

Qy 355 LeuAspIleSerAlaGluPro-----TrpLeuThrPhe 365

Db 2682 GCGTTGATTTCTTCGCCACCATAAATGAACGATTTTCATCCACTGTTGTTTGTACTTTT 2623

Qy 366 GlyValGlnProAsnProProValProSerProLysLeuGlyTrpMetAspGluAlaMet 385

Db 2622 TCTGTGTGTTACCTTCGCCCAACTTTTCCCTGTTTAATGGGTTCTTAGTGTGTTGTT 2563

Qy 386 AlaIle-----AspProPheAsnSerAspArgMetLeuTyrGlyThr--- 399

Db 2562 GTAATGTTTGTTCCTGGTTCACCTTTTGAACGACTTCTCTGTACCTGCGCTAAG 2503

Qy 400 -----GlyAlaThrLeuTyrAlaThrAsnAspLeuThrLysTrpAsp----- 413

Db 2502 TTTGGATCAAAATCGGTTTTTTATCAACGGAAATTTCTTCCGTTGACGTAAATCGGATCT 2443



Search completed: March 2, 2006, 13:23:14  
Job time : 927.098 secs

```
Qy 414 ---SerGlyGlyGlnIleHisIleAlaProMetValIysGlyLeuGluGluThrAlaVal 432
Db 2442 CCATCAACTGGACCATATTTTGTACATCATCATCTGGTGGAGTAACCTTCCGCTGTA 2383
Qy 433 AsnAspLeuIleSerPro-----ProSerGlyAla--- 442
Db 2382 TCAGGATTTTAACTCTCTGGTTTACCTGGAAACGCTCTTGGTACCTTTCGGTGCAATT 2323
Qy 443 -----ProLeuIleSerAlaLeuGly 449
Db 2322 GGATCAAAATTCATCTTATGCGCTGGCTTGATTTCTTCGCCACCATATTTCTGTGATTCA 2263
Qy 450 AspLeuGlyGlyPheThrHisAlaAspValThrAlaValProSerThrIlePheThrSer 469
Db 2262 TCTACTGGTGTGTTTGTGTACTTTTCTGTGTTTCA--CCTTCGCCAACTTTT--TCC 2209
Qy 470 ProVal-----PheThrThrGlyThrSerValAspTyrAlaGluLeuAsnProSer 486
Db 2208 CCTGTTAATGGGTCTTAGTTGTTGGTGTGTAATTGTTTGTCTCTGGTTCACCTTTC 2149
Qy 487 IleIleValArgAla----- 491
Db 2148 TGTTTAACGCGCTCTTACCTGGTTTAAATCAGATGAATTCAAGTTTCTTGTGCGAAT 2089
Qy 492 -----GlySerPheAsp-----ProSerSerGlnProAsnAspArgHisVal 505
Db 2088 GGAATTTCTTCGGTTCAGCTGATCGGATCTCCATCAACTGGACCA----- 2044
Qy 506 AlaPheSerThrAspGlyGlyLysAsnTrpPheGlnGlySerGluProGlyGlyValThr 525
Db 2043 -----TATTTTGTCAATCATCCACTGCTGGTGGTGTGACT 2011
Qy 526 ThrGlyGlyThrValAlaAlaSerAlaAspGlySerArgPheValTrrAlaProGlyAsp 545
Db 2010 ACTTCGCTGTA-----TCAGGGTTTITTA---ACTCCTGGTITA 1975
Qy 546 ProGlyGlnProValValTyrAlaValGlyPheGlyAsnSerTrpAlaAlaSerGlnGly 565
Db 1974 CCTGGAACG-----TCCCTTGGCTACCTTTCGGTGCA 1942
Qy 566 ValProAlaAsnAlaGlnIleArgSerAspArgValAsnProLysThrPheTyrAlaLeu 585
Db 1941 TTTGGATCAAAATTCATCCTTATGCTGCTGGCTTGATTTCTTCGCCACCATATTTCTGTGATT 1882
Qy 586 SerAsn---GlyThrPheTyrArgSerThrAspGly-----GlyValThrPheGlnPro 602
Db 1881 TCATCTACTGTTGTTTGTGTTTATTTTCTGTGTTGGTTCACTTCGCCAACCTTTTCCCT 1822
Qy 603 ValAlaAlaGlyLeuProSerSerGlyAlaValGlyValMetPheHisAlaValProGly 622
Db 1821 GTTAATGGTTCITA-----GTTGTTGGTGTGTTAATGTTTGTTCCTGGT 1774
Qy 623 LysGluGlyAspLeuTrpLeuAlaAlaSerSerGlyLeuTyrHisSerThrAsnGlyGly 642
Db 1773 -----TCACCTTTTGTGTTTAAACAGCTCTTCACCTGGTTTAAATCA-----GGATTG 1726
Qy 643 SerSerTrpSerAlaIleThrGlyValSerSerAlaValAsnValGlyPheGlyLysSer 662
Db 1725 AATTCACGTTTCTGTGTCGAATGGAATTTCTTCGTTGACGTGATCGGATCTCCATCAACT 1666
Qy 663 AlaProGlySerSerTyrProAlaValPheValValGlyThrIleGlyGlyValThrGly 682
Db 1665 GGACCA-----TATTTTGTCAATCATCCACAGGTGGAGTAACCTACT 1624
Qy 683 AlaTyrArgSerAsp-----AspCysGlyThrThrTrp 693
Db 1623 TCGCCTGTATCAGGATTTTAAACCCCGGCTTACCTGGTTGCGTGTGTTGACTACCTTT 1564
Qy 694 ValLeuIleAsnAspAspGlnHisGlnTyrGly 704
Db 1563 CGGTGCATTTG-GATCAAAATTCATCCTTATGGC 1532
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 13:55:49 ; Search time 5.12914 Seconds  
(without alignments)  
1434.576 Million cell updates/sec

Title: US-09-917-376-4  
Perfect score: 469  
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA\*  
1: /cgn2\_6/ptodata/1/iaa/5 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description       |
|------------|-------|-------------|--------|-------|-------------------|
| 1          | 195.5 | 41.7        | 616    | 2     | US-09-136-574A-47 |
| 2          | 194.5 | 41.5        | 1751   | 2     | US-09-136-574A-44 |
| 3          | 191.5 | 40.8        | 1426   | 2     | US-09-136-574A-43 |
| 4          | 172   | 36.7        | 700    | 1     | US-07-862-588B-2  |
| 5          | 165.5 | 35.3        | 551    | 1     | US-09-033-537A-1  |
| 6          | 165   | 35.2        | 167    | 4     | PCT-US95-13813-9  |
| 7          | 165   | 35.2        | 476    | 2     | US-09-339-159B-4  |
| 8          | 165   | 35.2        | 493    | 2     | US-09-198-956-10  |
| 9          | 165   | 35.2        | 493    | 2     | US-09-198-955A-12 |
| 10         | 165   | 35.2        | 493    | 2     | US-09-694-531-12  |
| 11         | 165   | 35.2        | 493    | 2     | US-09-670-141-10  |
| 12         | 165   | 35.2        | 493    | 2     | US-10-072-152-12  |
| 13         | 155.5 | 33.2        | 1352   | 2     | US-09-784-554B-2  |
| 14         | 150.5 | 32.1        | 1350   | 2     | US-09-784-554B-4  |
| 15         | 120.5 | 25.7        | 531    | 1     | US-07-862-588B-7  |
| 16         | 114   | 24.3        | 162    | 1     | US-08-048-164A-2  |
| 17         | 114   | 24.3        | 162    | 1     | US-08-460-462-2   |
| 18         | 114   | 24.3        | 162    | 1     | US-08-460-457-2   |
| 19         | 114   | 24.3        | 162    | 1     | US-08-460-458-2   |
| 20         | 114   | 24.3        | 162    | 1     | US-08-460-455-2   |
| 21         | 114   | 24.3        | 162    | 1     | US-08-330-394A-2  |
| 22         | 114   | 24.3        | 163    | 2     | US-09-006-636-7   |
| 23         | 114   | 24.3        | 163    | 2     | US-09-006-632-7   |
| 24         | 114   | 24.3        | 163    | 2     | US-09-325-274-7   |
| 25         | 113   | 24.1        | 382    | 2     | US-09-277-716-22  |
| 26         | 113   | 24.1        | 382    | 2     | US-09-609-161B-22 |
| 27         | 112   | 23.9        | 154    | 1     | US-08-330-394A-29 |

|    |      |      |      |   |                      |                   |
|----|------|------|------|---|----------------------|-------------------|
| 28 | 112  | 23.9 | 156  | 1 | US-08-330-394A-22    | Sequence 22, Appl |
| 29 | 72.5 | 15.5 | 272  | 2 | US-09-902-540-11056  | Sequence 11056, A |
| 30 | 68.5 | 14.6 | 500  | 2 | US-10-104-047-2825   | Sequence 2825, Ap |
| 31 | 65.5 | 14.0 | 1566 | 2 | US-09-581-472B-2     | Sequence 2, Appli |
| 32 | 65   | 13.9 | 474  | 2 | US-09-248-796A-25524 | Sequence 25524, A |
| 33 | 64   | 13.6 | 143  | 2 | US-09-301-593-26     | Sequence 26, Appl |
| 34 | 64   | 13.6 | 428  | 2 | US-09-118-319-5      | Sequence 5, Appli |
| 35 | 64   | 13.6 | 464  | 1 | US-08-353-400-36     | Sequence 36, Appl |
| 36 | 64   | 13.6 | 472  | 2 | US-09-301-593-30     | Sequence 30, Appl |
| 37 | 63.5 | 13.5 | 1785 | 2 | US-09-341-587-3      | Sequence 3, Appli |
| 38 | 63   | 13.4 | 453  | 2 | US-09-301-593-18     | Sequence 18, Appl |
| 39 | 63   | 13.4 | 472  | 2 | US-09-301-593-43     | Sequence 43, Appl |
| 40 | 63   | 13.4 | 718  | 2 | US-09-328-352-4640   | Sequence 4640, Ap |
| 41 | 63   | 13.4 | 1133 | 2 | US-09-902-540-12243  | Sequence 12243, A |
| 42 | 63   | 13.4 | 1581 | 2 | US-09-110-517-2      | Sequence 2, Appli |
| 43 | 62.5 | 13.3 | 288  | 2 | US-09-423-439-38     | Sequence 38, Appl |
| 44 | 62.5 | 13.3 | 389  | 2 | US-09-902-540-14347  | Sequence 14347, A |
| 45 | 62.5 | 13.3 | 445  | 1 | US-08-353-400-33     | Sequence 33, Appl |

## ALIGNMENTS

## RESULT 1

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 616 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 2; Length 616;  
Best Local Similarity 43.2%; Pred. No. 3.4e-13;  
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

|    |    |                                     |                      |    |
|----|----|-------------------------------------|----------------------|----|
| Qy | 1  | VSGGVKQYKNDSPAGDNOIKPGIQLVNTGSSSDLS | STVTVRVWFTPDGSSFLVNC | 60 |
| Db | 1  | MGSGVKVLYKNNETASTGSRIPWPKVNGSSVDLS  | SRVKIRYWTVDGKPOS     | 59 |
| Qy | 61 | WAAMGCGNIRASFGSVNPATPTADTYLQ        | 88                   |    |
| Db | 60 | WAQIGASNVTFNFVKLSGSGVSGADYYLE       | 87                   |    |

## RESULT 2

US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Berquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

Williams, Diane F.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; ;  
; ;

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>

```

, INFORMATION FOR SEQ ID NO: 44:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1751 amino acids
, TYPE: amino acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, MOLECULE TYPE: protein
, SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-136-574A-44

```

Query Match 41.5%; Score 194.5; DB 2; Length 1751;  
Best Local Similarity 44.7%; Pred. No. 1.6e-12;  
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSAPGDNQIKPGLQLVNTGSSSSVDLSTVTVRYWFTRDGSSTLYVNCDAW 63

|            |     |  |     |
|------------|-----|--|-----|
| D <b>b</b> | 678 | GKVLKYKNNETSASTGSIRPWFKIVGGSSVDLSRVKIRYWTVDGDKQSAV-CDWAQ | 736 |
| Q <b>y</b> | 64  | MCCGNIRASFGSVNPA <del>T</del> PADTYLQ                    | 88  |
| D <b>b</b> | 737 | IGASNVTNFVKLLSGSGVGDYLYE                                 | 761 |

### RESULTS

US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

**TITLE OF INVENTION:** Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESS: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:

```

; name: Bob, Mary E.
;
; REGISTRATION NUMBER: 31,215
; REFERENCE/SOCKET NUMBER: 1997US001/CIP
;
; TELECOMMUNICATION INFORMATION:
;
;   TELEPHONE: 215-540-9200
;   TELEFAX: 215-540-5818
;
; TELEX: <Unknown>

```

```

, SEQA. COMMENTS
, INFORMATION FOR SEQ ID NO: 43:
, =====
, SEQUENCE CHARACTERISTICS:
,     LENGTH: 1426 amino acids
,     TYPE: amino acid
,     STRANDEDNESS: single
,     TOPOLOGY: linear
,     MOLECULE TYPE: N0. 6294366e
,     SEQUENCE DESCRIPTION: SEQ ID
,     US-09-136-574A-43

```

```

Query Match      40.8%; Score 191.5; DB 2; Length 1426;
Best Local Similarity 43.5%; Pred. No. 2.6e-12;
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQVKNDSDAPGDNQIKPGLQVLNTGSSSDLSITVTVYWFTRDGSSTLVVNCDDAA 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 GKKVYNNSTACTSGISPPKPIVYVNCSSSSVNI SVKIDYKTVYVNCDDVCSAV-CDDBO 471

```

|    |     |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |    |   |   |   |    |   |
|----|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|----|---|
| Qy | 4   | G | V | K | V | Q | K | N | D | S | A | P | G | N | I | K | P | L | Q | L | V | N | T | G | S | S | V | D | L | T | T | V | R | Y | F | W | F | T | R | D | G | S | S | T | L | V | N | C | D | W | A | 63 |   |   |   |    |   |
|    |     | - | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  | : | : | : | :  | : |
| Db | 413 | G | L | K | V | L | Y | K | N | N | E | T | S | A | T | G | S | I | R | P | F | K | I | V | N | G | G | S | S | V | D | L | S | R | K | I | Y | W | T | V | D | G | D | K | P | O | S | A | V | - | C | D  | W | A | Q | 47 |   |

QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88  
: | | : | : : | | :  
Db 472 IGASNVTNFVKLSGGVSCADYYLE 496

#### RESULT, T 4

```

US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2

Query Match 36.7%; Score 172; DB 1; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.5e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3

QY 1 VSGGVKQVQYKND-SAPGDNIKPGLOLVNTGSSVLDLSTVTYVYFTRDGGSSTLVYNC 59
Db 549 VNSDLVVQVKGRNATDNIKPHFNIQNGKTSFVLDLSLTURYFTKD-SSAAMNGWI 607
QY 60 DWAAAGCGNIRASFGSVNPAFTPTADTYLQ 88
Db 608 DWAKLGGSNIQISFGNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59580830 No. 5958083disk of No. 5958083th America, Inc.

```



```

; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
; US-09-670-141-10

Query Match 35.2%; Score 165; DB 2; Length 493;
Best Local Similarity 36.1%; Pred. No. 5.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

QY 1 VSGGVKQVQYKNDSDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWTFTRDGGSTLVNCD 60
Db 331 VSGNLKVEFVNSPDTTNSINPQFKVTNTGSSAIDLKLTLLYYITVDGQKQTFW-CD 389
QY 61 WAAM-----GCCNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIIGSGSYNGITSNVRKGTFFVKMSSSTNNADTYLE 426

RESULT 12
US-10-072-152-12
; Sequence 12, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match
Best Local Similarity 35.2%; Score 165; DB 2; Length 493;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSSTLVNCD 60
Db 331 VSGNLKVEFYNSPDSITNSINPQFKVNTGSSAIDLSKULFRIYYTVGQKQOTFW-CD 389

Qy 61 WAAM-----CGGNIRASFGSVNPATPTADTYLQ 88
Db 390 HAAIGSGSYNGITSNVKGFVKMSSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784.554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match
Best Local Similarity 33.2%; Score 155.5; DB 2; Length 1352;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1200 GLLLQYRTADTKYNDHNLNPFQVLVNGTTSIPINELKIRYYITDGDREQ-TFNCDYAT 1258

Qy 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 1259 LSCSKLNGKLVKMDKATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784.554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match
Best Local Similarity 32.1%; Score 150.5; DB 2; Length 1350;
```

```
Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDSPAGDNOIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1199 GLVQYRTADTKYNDHNLNPFQVLVNGTTSIPINELKIRYYITDGDREQ-TFNCDYAV 1257

Qy 64 MCGGNIRASFGSVNPATPTADTYLQ 88
Db 1258 LSCSKLNGKLVKMDKATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match
Best Local Similarity 25.7%; Score 120.5; DB 1; Length 531;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

Qy 2 SGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSLTVTVRYWFTRDGSSSTLVNCDW 61
Db 432 TGNLVQYKVGDTSATDNQKPSFNKNGTTPVNLGLKXXXXKD-GPADMSCSIDW 490

Qy 62 AAMCGGNIRASFGSVNPATPTADTY 86
Db 491 AQIGRTNVLAF--ANFTGSNTDTY 513

Search completed: March 2, 2006, 13:58:00
```



Job time : 6.12914 secs

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:20:38 ; Search time 17.1446 Seconds  
(without alignments)  
2169.009 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 469

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description         |
|------------|-------|-------------|--------|-------|---------------------|
| 1          | 469   | 100.0       | 88     | 3     | US-09-917-376-5     |
| 2          | 469   | 100.0       | 88     | 4     | US-10-155-400-5     |
| 3          | 469   | 100.0       | 89     | 3     | US-09-917-376-4     |
| 4          | 469   | 100.0       | 89     | 4     | US-10-155-400-4     |
| 5          | 469   | 100.0       | 154    | 3     | US-09-917-378-4     |
| 6          | 469   | 100.0       | 762    | 3     | US-09-917-378-1     |
| 7          | 466   | 99.4        | 150    | 3     | US-09-917-384-5     |
| 8          | 466   | 99.4        | 150    | 3     | US-09-917-383-5     |
| 9          | 466   | 99.4        | 1043   | 3     | US-09-917-384-6     |
| 10         | 466   | 99.4        | 1043   | 3     | US-09-917-383-6     |
| 11         | 466   | 99.4        | 1228   | 3     | US-09-917-384-1     |
| 12         | 466   | 99.4        | 1228   | 3     | US-09-917-383-1     |
| 13         | 462   | 98.5        | 957    | 3     | US-09-917-376-1     |
| 14         | 462   | 98.5        | 957    | 4     | US-10-155-400-1     |
| 15         | 257   | 54.8        | 741    | 4     | US-10-156-761-8100  |
| 16         | 190   | 40.5        | 508    | 4     | US-10-369-493-23151 |
| 17         | 188   | 40.1        | 477    | 4     | US-10-466-208-12    |
| 18         | 188   | 40.1        | 496    | 4     | US-10-466-208-8     |
| 19         | 188   | 40.1        | 677    | 4     | US-10-433-577-35    |
| 20         | 180   | 38.4        | 1621   | 4     | US-10-185-990-10    |
| 21         | 165   | 35.2        | 167    | 5     | US-10-933-404-4     |
| 22         | 165   | 35.2        | 476    | 4     | US-10-372-054-4     |
| 23         | 165   | 35.2        | 493    | 4     | US-10-072-152-12    |
| 24         | 165   | 35.2        | 493    | 4     | US-10-655-433-12    |
| 25         | 165   | 35.2        | 599    | 3     | US-09-955-555A-29   |
| 26         | 155.5 | 33.2        | 1352   | 3     | US-09-784-554B-2    |
| 27         | 155.5 | 33.2        | 1352   | 5     | US-10-896-555-2     |

|    |       |      |      |   |                      |                    |
|----|-------|------|------|---|----------------------|--------------------|
| 28 | 150.5 | 32.1 | 1350 | 3 | US-09-784-554B-4     | Sequence 4, Appli  |
| 29 | 150.5 | 32.1 | 1350 | 5 | US-10-896-555-4      | Sequence 4, Appli  |
| 30 | 130.5 | 27.8 | 1483 | 4 | US-10-282-122A-51483 | Sequence 51483, A  |
| 31 | 114   | 24.3 | 256  | 4 | US-10-261-446-6      | Sequence 6, Appli  |
| 32 | 114   | 24.3 | 256  | 4 | US-10-261-446B-6     | Sequence 6, Appli  |
| 33 | 114   | 24.3 | 256  | 5 | US-10-782-234-6      | Sequence 6, Appli  |
| 34 | 113   | 24.1 | 382  | 3 | US-09-808-898-22     | Sequence 22, Appli |
| 35 | 105.5 | 22.5 | 163  | 4 | US-10-460-524-2      | Sequence 2, Appli  |
| 36 | 78.5  | 16.7 | 1049 | 4 | US-10-282-122A-49900 | Sequence 49900, A  |
| 37 | 75.5  | 16.1 | 346  | 4 | US-10-437-963-179234 | Sequence 179234,   |
| 38 | 75.5  | 16.1 | 618  | 4 | US-10-211-462-223    | Sequence 223, App  |
| 39 | 75.5  | 16.1 | 618  | 5 | US-10-723-860-4039   | Sequence 4039, Ap  |
| 40 | 75.5  | 16.1 | 618  | 5 | US-10-756-149-5713   | Sequence 5713, Ap  |
| 41 | 75.5  | 16.1 | 986  | 3 | US-09-747-835A-53    | Sequence 53, Appl  |
| 42 | 75.5  | 16.1 | 986  | 4 | US-10-120-604-101    | Sequence 101, App  |
| 43 | 75.5  | 16.1 | 986  | 4 | US-10-225-567A-406   | Sequence 406, App  |
| 44 | 75.5  | 16.1 | 986  | 4 | US-10-398-458-3      | Sequence 3, Appli  |
| 45 | 75.5  | 16.1 | 986  | 4 | US-10-312-312-53     | Sequence 53, Appl  |

## ALIGNMENTS

### RESULT 1

US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.40S01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

|                       |        |  |      |                                 |
|-----------------------|--------|--|------|---------------------------------|
| Query Match           | 100.0% | Score 469  | DB 3 | Length 88                       |
| Best Local Similarity | 100.0% | Pred. No. 3.6e-48  |      |                                 |
| Matches               | 88     | Conservative   | 0    | Mismatches 0; Indels 0; Gaps 0; |
| QY                    | 1      | VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTVTYVWFRDGGSSTLVYVNC | 60   |                                 |
| DB                    | 1      | VSGGVKQVKNDSAPGDNQIKPGLQLVNTGSSVDLSVTVTYVWFRDGGSSTLVYVNC | 60   |                                 |
| QY                    | 61     | WAAMGCGNIRASFGSVNPATPTADTYLQ                             | 88   |                                 |
| DB                    | 61     | WAAMGCGNIRASFGSVNPATPTADTYLQ                             | 88   |                                 |

### RESULT 2

US-10-155-400-5  
; Sequence 5, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400

; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosoliticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-5

Query Match 100.0%; Score 469; DB 4; Length 88;  
Best Local Similarity 100.0%; Pred. No. 3.7e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 3  
US-09-917-376-4  
; Sequence 4, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosoliticus  
; NAME/KEY: MOD RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-4

Query Match 100.0%; Score 469; DB 3; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.7e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 4  
US-10-155-400-4  
; Sequence 4, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOGLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 89  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulosoliticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (89)  
; OTHER INFORMATION: Any amino acid  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-10-155-400-4

Query Match 100.0%; Score 469; DB 4; Length 89;  
Best Local Similarity 100.0%; Pred. No. 3.7e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 5  
US-09-917-378-4  
; Sequence 4, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 154  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate  
; OTHER INFORMATION: binding domain  
US-09-917-378-4

Query Match 100.0%; Score 469; DB 3; Length 154;  
Best Local Similarity 100.0%; Pred. No. 7.1e-48;  
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 1 VSGGVKQYKNDSPAGDNOIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 6  
US-09-917-378-1

Sequence 1, Application US/09917378  
 Publication No. US20030119093A1  
 GENERAL INFORMATION:  
 APPLICANT: DING, SHI-YOU  
 APPLICANT: ADNEY, WILLIAM S.  
 APPLICANT: VINZANT, TODD B.  
 APPLICANT: DECKER, STEPHEN R.  
 APPLICANT: HIMMEL, MICHAEL E.  
 TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
 TITLE OF INVENTION: CELLULOXYTICUS  
 FILE REFERENCE: 40197.7US01  
 CURRENT APPLICATION NUMBER: US/09/917,378  
 CURRENT FILING DATE: 2001-07-28  
 NUMBER OF SEQ ID NOS: 8  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 1  
 LENGTH: 762  
 TYPE: PRT  
 ORGANISM: Acidothermus cellulolyticus  
 US-09-917-378-1

Query Match 100.0%; Score 469; DB 3; Length 762;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-47;  
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 DB 455 VSGGVKQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 514  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542

## RESULT 7

US-09-917-384-5  
 Sequence 5, Application US/09917384  
 Publication No. US20030096342A1  
 GENERAL INFORMATION:  
 APPLICANT: DING, SHI-YOU  
 APPLICANT: ADNEY, WILLIAM S.  
 APPLICANT: VINZANT, TODD B.  
 APPLICANT: DECKER, STEPHEN R.  
 APPLICANT: HIMMEL, MICHAEL E.  
 TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 TITLE OF INVENTION: CELLULOXYTICUS  
 FILE REFERENCE: 40170.6US01  
 CURRENT APPLICATION NUMBER: US/09/917,384  
 CURRENT FILING DATE: 2001-07-28  
 NUMBER OF SEQ ID NOS: 14  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 5  
 LENGTH: 150  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 OTHER INFORMATION: Guxa  
 US-09-917-384-5

Query Match 99.4%; Score 466; DB 3; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 1.6e-47;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 DB 1 VSGGLKVQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 8

US-09-917-383-5  
 Sequence 5, Application US/09917383  
 Publication No. US20030104522A1  
 GENERAL INFORMATION:  
 APPLICANT: DING, SHI-YOU  
 APPLICANT: ADNEY, WILLIAM S.  
 APPLICANT: VINZANT, TODD B.  
 APPLICANT: DECKER, STEPHEN R.  
 APPLICANT: HIMMEL, MICHAEL E.  
 TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 TITLE OF INVENTION: CELLULOXYTICUS  
 FILE REFERENCE: 40170.6US01  
 CURRENT APPLICATION NUMBER: US/09/917,383  
 CURRENT FILING DATE: 2001-07-28  
 NUMBER OF SEQ ID NOS: 14  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 5  
 LENGTH: 150  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 OTHER INFORMATION: Guxa  
 US-09-917-383-5

Query Match 99.4%; Score 466; DB 3; Length 150;  
 Best Local Similarity 98.9%; Pred. No. 1.6e-47;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 DB 1 VSGGLKVQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
 DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

## RESULT 9

US-09-917-384-6  
 Sequence 6, Application US/09917384  
 Publication No. US20030096342A1  
 GENERAL INFORMATION:  
 APPLICANT: DING, SHI-YOU  
 APPLICANT: ADNEY, WILLIAM S.  
 APPLICANT: VINZANT, TODD B.  
 APPLICANT: DECKER, STEPHEN R.  
 APPLICANT: HIMMEL, MICHAEL E.  
 TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
 TITLE OF INVENTION: CELLULOXYTICUS  
 FILE REFERENCE: 40170.6US01  
 CURRENT APPLICATION NUMBER: US/09/917,384  
 CURRENT FILING DATE: 2001-07-28  
 NUMBER OF SEQ ID NOS: 14  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 6  
 LENGTH: 1043  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: Segment of  
 OTHER INFORMATION: Guxa  
 US-09-917-384-6

Query Match 99.4%; Score 466; DB 3; Length 1043;  
 Best Local Similarity 98.9%; Pred. No. 1.7e-46;  
 Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
 DB 477 VSGGLKVQYKND SAPGNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 536

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

RESULT 10  
US-09-917-383-6  
; Sequence 6, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of Guxa  
US-09-917-383-6

Query Match 99.4%; Score 466; DB 3; Length 1043;  
Best Local Similarity 98.9%; Pred. No. 1.7e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 477 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 536

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

RESULT 11  
US-09-917-384-1  
; Sequence 1, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of Guxa  
US-09-917-384-1

Query Match 99.4%; Score 466; DB 3; Length 1228;  
Best Local Similarity 98.9%; Pred. No. 2e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 584 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 643

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 12  
US-09-917-383-1  
; Sequence 1, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of Guxa  
US-09-917-383-1

Query Match 99.4%; Score 466; DB 3; Length 1228;  
Best Local Similarity 98.9%; Pred. No. 2e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VSGGVKQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 60  
Db 584 VSGGLKVQYKNDSPAGDNQIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSSSTLVNCD 643

Qy 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
Db 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

RESULT 13  
US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid

US-09-917-376-1

Query Match 98.5%; Score 462; DB 3; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.5e-46;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 60  
 Db 869 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 60  
 QY 61 WAAAGCGNIRASFGSVNPATPTADTYLQ 88  
 Db 929 WAAAGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 14

US-10-155-400-1  
 ; Sequence 1, Application US/10155400  
 ; Publication No. US20030108988A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DING, SHI-YOU  
 ; APPLICANT: ADNEY, WILLIAM S.  
 ; APPLICANT: VINZANT, TODD B.  
 ; APPLICANT: HIMMEL, MICHAEL E.  
 ; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
 ; TITLE OF INVENTION: CELLULOXYTICUS  
 ; FILE REFERENCE: NREL 01-36A  
 ; CURRENT APPLICATION NUMBER: US/10/155,400  
 ; CURRENT FILING DATE: 2002-10-22  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 957  
 ; TYPE: PRT  
 ; ORGANISM: Acidothermus cellulolyticus  
 ; FEATURE:  
 ; NAME/KEY: MOD\_RES  
 ; LOCATION: (957)  
 ; OTHER INFORMATION: Any amino acid  
 US-10-155-400-1

Query Match 98.5%; Score 462; DB 4; Length 957;  
 Best Local Similarity 97.7%; Pred. No. 4.5e-46;  
 Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 60  
 Db 869 VSGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 60  
 QY 61 WAAAGCGNIRASFGSVNPATPTADTYLQ 88  
 Db 929 WAAAGCGNIRASFGSVNPATPTADTYLQ 956

RESULT 15

US-10-156-761-8100  
 ; Sequence 8100, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 8100  
 ; LENGTH: 741  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-8100

Query Match 54.8%; Score 257; DB 4; Length 741;  
 Best Local Similarity 56.3%; Pred. No. 1.2e-21;  
 Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;  
 QY 2 SGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 61  
 Db 591 SGGVKVOYKNDSPAGDNQIKPGLQVNTGSSVDLSVTYVYFTRDGGSSSTLYVNC 61  
 QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88  
 Db 651 AAMGCGNIRASFGSVNPATPTADTYLQ 677

Search completed: March 2, 2006, 14:26:55  
 Job time: 18.1446 secs

| Result No. | Score | Match | Length | ID                  | Description        |
|------------|-------|-------|--------|---------------------|--------------------|
| 1          | 163   | 34.8  | 499    | US-10-517-939-328   | Sequence 328, Appl |
| 2          | 157   | 33.5  | 534    | US-10-510-386-230   | Sequence 230, Appl |
| 3          | 113   | 24.1  | 382    | US-11-179-411-22    | Sequence 22, Appl  |
| 4          | 113   | 24.1  | 382    | US-11-175-766-22    | Sequence 22, Appl  |
| 5          | 108   | 23.0  | 738    | US-10-517-939-344   | Sequence 344, Appl |
| 6          | 72    | 15.4  | 210    | US-11-170-653-51    | Sequence 51, Appl  |
| 7          | 70    | 14.9  | 5291   | US-11-052-554A-281  | Sequence 281, Appl |
| 8          | 68.5  | 14.6  | 500    | US-11-072-512-2825  | Sequence 2825, Ap  |
| 9          | 68    | 14.5  | 755    | US-10-517-939-330   | Sequence 330, Appl |
| 10         | 64.5  | 13.8  | 389    | US-11-207-626A-26   | Sequence 26, Appl  |
| 11         | 62    | 13.2  | 389    | US-11-207-636A-20   | Sequence 20, Appl  |
| 12         | 61    | 13.0  | 350    | US-10-517-939-14    | Sequence 14, Appl  |
| 13         | 61    | 13.0  | 1889   | US-11-102-476-46    | Sequence 46, Appl  |
| 14         | 60.5  | 12.9  | 806    | US-11-098-686-10278 | Sequence 10278, A  |
| 15         | 60.5  | 12.9  | 983    | US-11-087-099-7483  | Sequence 7483, Ap  |
| 16         | 60    | 12.8  | 86     | US-11-102-476-24    | Sequence 24, Appl  |
| 17         | 60    | 12.8  | 91     | US-11-102-476-9     | Sequence 9, Appli  |
| 18         | 60    | 12.8  | 919    | US-11-074-176-284   | Sequence 284, App  |
| 19         | 60    | 12.8  | 1225   | US-11-102-476-2     | Sequence 2, Appli  |
| 20         | 60    | 12.8  | 1424   | US-11-102-476-4     | Sequence 4, Appli  |
| 21         | 59.5  | 12.7  | 217    | US-10-454-437-82    | Sequence 82, Appl  |
| 22         | 59    | 12.6  | 111    | US-11-072-512-2820  | Sequence 2820, Ap  |
| 23         | 59    | 12.6  | 124    | US-10-771-257-19    | Sequence 19, Appl  |
| 24         | 59    | 12.6  | 124    | US-11-127-677-19    | Sequence 19, Appl  |
| 25         | 59    | 12.6  | 182    | US-11-170-653-49    | Sequence 49, Appl  |





; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 344  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
US-10-517-939-344

Query Match 23.0%; Score 108; DB 6; Length 738;  
Best Local Similarity 32.1%; Pred. No. 5.2e-05;  
Matches 27; Conservative 18; Mismatches 31; Indels 8; Gaps 4;  
Qy 7 VQYKNDSPGDN--QIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSGSTLVYNCDAAM 64  
Db 591 LQYR---SADGNNYQMKPQETIKNAGKVPILSELIRIYVTFPE-STQPDVTRIDWAQF 646  
Qy 65 GCGNIRASFGSVNPTATPDATYQLQ 88  
Db 647 GAHVQVTV--VPPSAAAHAAYVE 668

RESULT 6  
US-11-170-653-51  
; Sequence 51, Application US/11/170653  
; Publication No. US20050271769A1  
; GENERAL INFORMATION:  
; APPLICANT: Danisco A/S  
; APPLICANT: Sibbesen, Ole  
; APPLICANT: Sorensen, Jens  
; TITLE OF INVENTION: Xylanase Variants Having Altered Sensitivity to Xylanase Inhibitor  
; FILE REFERENCE: 674509-2046  
; CURRENT APPLICATION NUMBER: US/11/170,653  
; CURRENT FILING DATE: 2005-06-23  
; PRIOR APPLICATION NUMBER: US/10/237,386  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: PCT/IB01/00426  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: GB 0005585.5  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: GB 0015751.1  
; PRIOR FILING DATE: 2000-06-27  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 210  
; TYPE: PRT  
; ORGANISM: A. tubigenis  
US-11-170-653-51

Query Match 15.4%; Score 72; DB 7; Length 210;  
Best Local Similarity 31.5%; Pred. No. 0.23;  
Matches 23; Conservative 10; Mismatches 22; Indels 18; Gaps 4;  
Qy 29 NTGSSVDLSTVTV-RVWFTFDG-----GSSLVYNCDAAMCGNIRASFG 74  
Db 40 NLGDFTYDESAGTFSMYW--EDGVSDFFVGLWGTGSSITTSYSAEYASGSASGLAVYG 97  
Qy 75 SVNPTATPDATYL 87  
Db 98 WVN--YQAEYII 108

RESULT 7  
US-11-052-554A-281

; Sequence 281, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 281  
; LENGTH: 5291  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-281

Query Match 14.9%; Score 70; DB 7; Length 5291;  
Best Local Similarity 33.3%; Pred. No. 18;  
Matches 29; Conservative 15; Mismatches 19; Indels 24; Gaps 7;  
Qy 17 GD---NQIKPGLQLVNTGSS--VDL-STVTVRYWFTRDG---SSTLVYNCDAW----- 61  
Db 848 GDDVINAIVEKGETLVVSGSTSGVEAGQTVTVTF-----GCKNYTTTVEANGSWTVNVPPA 902  
Qy 62 --AM--GCGNIRASFGSVNPTATPD 84  
Db 903 DLALPDGAGNVQASVNSNGNSAQAD 929

RESULT 8  
US-11-072-512-2825  
; Sequence 2825, Application US/11072512  
; Publication No. US20060029945A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOUYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: 084335-0191  
; CURRENT APPLICATION NUMBER: US/11/072,512  
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/350,978  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: JP 2001-379298  
; PRIOR FILING DATE: 2001-11-05  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2825  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-072-512-2825

Query Match 14.6%; Score 68.5; DB 7; Length 500;  
Best Local Similarity 33.3%; Pred. No. 1.7; Indels 11; Gaps 3;  
Matches 19; Conservative 9; Mismatches 18; Indels 11; Gaps 3;

QY 31 GSSVDLSTVTVRYWFTRDG-----GSS--TLVNC-----WAAMGCGNIRASFGSV 76  
DB 414 GTPETDIDSSCSRYTLKADGQCPSGSGTIVITCEPISAYGARGSANIKVTISV 470

RESULT 9  
US-10-517-939-330  
; Sequence 330, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Blum, David  
; APPLICANT: Esteghlalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 2578-4123.2US  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 330  
; LENGTH: 755  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
; NAME/KEY: SIGNAL  
; LOCATION: (1)....(35)  
US-10-517-939-330

Query Match 14.5%; Score 68; DB 6; Length 755;  
Best Local Similarity 30.1%; Pred. No. 3.1;  
Matches 31; Conservative 8; Mismatches 40; Indels 24; Gaps 5;

QY 2 SGVKVQYK-NDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWFTRD----- 49  
DB 651 SGACTVAYATNDWGS-----FTANVTLTNTGSA--LNGWTLAYAFPGNQTISNANGTA 705

QY 50 --GSSSTLVYNCDAAMCGNIRASFG-----SYNPATPTADT 85  
DB 706 VQSGSVSVTNAGWNSLPPNVASASFGQASYSNGNSVPSAFT 748

RESULT 10  
US-11-207-626A-26  
; Sequence 26, Application US/11207626A  
; Publication No. US20060014276A1  
; GENERAL INFORMATION:  
; APPLICANT: Havenga, Menzo  
; APPLICANT: Vogels, Ronald  
; APPLICANT: Bout, Abraham  
; TITLE OF INVENTION: CHIMERIC ADENOVIRUSES  
; FILE REFERENCE: 2578-4123.2US  
; CURRENT APPLICATION NUMBER: US/11/207,626A  
; CURRENT FILING DATE: 2005-08-18  
; PRIOR APPLICATION NUMBER: EP 98202297.2  
; PRIOR FILING DATE: 1998-07-08  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26

LENGTH: 389  
TYPE: PRT  
ORGANISM: Human Adenovirus 32 Fiber Protein  
US-11-207-626A-26

Query Match 13.8%; Score 64.5; DB 7; Length 389;  
Best Local Similarity 32.0%; Pred. No. 3.8;  
Matches 24; Conservative 10; Mismatches 32; Indels 9; Gaps 3;

QY 11 NDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCDAAMCGNIR 70  
DB 251 NNNTNPTDKKITVKL-LFNEKGVLMDSSSLKKEYWNYRDNDSSTQAY--DNAVPFMPNPK 307

QY 71 ASFGSVNPATPTADT 85  
DB 308 AY-----PKPTTDT 316

RESULT 11  
US-11-207-626A-20  
; Sequence 20, Application US/11207626A  
; Publication No. US20060014276A1  
; GENERAL INFORMATION:  
; APPLICANT: Havenga, Menzo  
; APPLICANT: Vogels, Ronald  
; APPLICANT: Bout, Abraham  
; TITLE OF INVENTION: CHIMERIC ADENOVIRUSES  
; FILE REFERENCE: 2578-4123.2US  
; CURRENT APPLICATION NUMBER: US/11/207,626A  
; CURRENT FILING DATE: 2005-08-18  
; PRIOR APPLICATION NUMBER: EP 98202297.2  
; PRIOR FILING DATE: 1998-07-08  
; NUMBER OF SEQ ID NOS: 87  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 20  
; LENGTH: 389  
; TYPE: PRT  
; ORGANISM: Human Adenovirus 24 Fiber Protein  
US-11-207-626A-20

Query Match 13.2%; Score 62; DB 7; Length 389;  
Best Local Similarity 33.3%; Pred. No. 7.5;  
Matches 25; Conservative 9; Mismatches 33; Indels 8; Gaps 3;

QY 11 NDSAPGDNQIKPGLQVLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCDAAMCGNIR 70  
DB 250 NNNTNPTDKKITVKL-LFNEKGVLMDSSSLKKEYWNYRDNDSSTQAY--DNAVPFMPNPK 307

QY 71 ASFGSVNPATPTADT 85  
DB 308 AY-----PKPTTDT 316

RESULT 12  
US-10-517-939-14  
; Sequence 14, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Blum, David  
; APPLICANT: Esteghlalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 2578-4123.2US  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299

```
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-517-939-14

Query Match      13.0%; Score 61; DB 6; Length 350;
Best Local Similarity 26.7%; Pred. No. 8.7;
Matches 16; Conservative 10; Mismatches 26; Indels 8; Gaps 1;

QY 10 KNDSPAGDNOIKPGLQVNTGSSVDLSTVTYVRYFTRDGGSTLTYNCDW 61
Db 40 RHYNLTADNQMKPESVLDRTATLAKGDLHLAAVDFTRVDALMYFARDNGIPMRVHTLAW 99

RESULT 13
US-11-102-476-46
; Sequence 46, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U.M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102,476
; CURRENT FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1889
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-46

Query Match      13.0%; Score 61; DB 7; Length 1889;
Best Local Similarity 26.2%; Pred. No. 62;
Matches 21; Conservative 14; Mismatches 27; Indels 18; Gaps 4;

QY 19 NOIKPGL-----QLVNTG-----SSVDLSTVTYVRYFTRDGGSTLTY-----NCDWA 62
Db 498 NPVNPFLAKGLTKFTATGIYSDNSKNDITSVTT--WFSDDSIATISNAQKNQGNAYGA 555

QY 63 AMGCNTRASFSGSVNPATPT 82
Db 556 ATGTTDIKATFGKVSVPST 575

RESULT 14
US-11-098-686-10278
; Sequence 10278, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10278
; LENGTH: 806
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10278

Query Match      12.9%; Score 60.5; DB 7; Length 806;
Best Local Similarity 33.3%; Pred. No. 26;
Matches 20; Conservative 5; Mismatches 20; Indels 15; Gaps 4;

QY 10 KNDSPAGDNOIKPGLQVNTGSSVDLSTVTYVRYWF--TRDG--GSSTLYV--NCDWAA 63
Db 286 KNNIVTPGKDPYPPSSQLVRNHQ-----RYWVRGMADGFIIGNSTWCYISNLDIVS 336

RESULT 15
US-11-087-099-7483
; Sequence 7483, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 7483
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. syringae B728a
US-11-087-099-7483

Query Match      12.9%; Score 60.5; DB 7; Length 983;
Best Local Similarity 33.3%; Pred. No. 33;
Matches 15; Conservative 9; Mismatches 10; Indels 11; Gaps 2;

QY 2 SGGVKVQYKNDNSAPGD--NQIKPGLQL-----VNTGSSSV 35
Db 522 NAGVKAIEFSNTSFSGDPVLTRVEFGVNLNWTGNSVTNAGSTAV 566

Search completed: March 2, 2006, 14:28:14
Job time : 2.8047 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:06:58 ; Search time 32.1046 Seconds  
(without alignments)  
4927.737 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 469

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table:

|                           |  |
|---------------------------|--|
| BLOSUM62                  |  |
| Xgapop 10.0 , Xgapext 0.5 |  |
| Ygapop 10.0 , Ygapext 0.5 |  |
| Fgapop 6.0 , Fgapext 7.0  |  |
| Delop 6.0 , Delext 7.0    |  |

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/abs/ABSSWEB\_spool/US09917376/runat\_02032006\_091454\_8101/app\_query.fasta\_1  
-DB=Issued\_Patents\_NA -QFWT=fastcap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abes06h -USER=US09917376 @CGN 1 278 @runat\_02032006\_091454\_8101  
-NCPU=6 -ICPU=3 -NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:

|    |                                       |
|----|---------------------------------------|
| 1: | /cgn2_6/ptodata/1/ina/1 COMB.seq:     |
| 2: | /cgn2_6/ptodata/1/ina/5 COMB.seq:     |
| 3: | /cgn2_6/ptodata/1/ina/6A COMB.seq:    |
| 4: | /cgn2_6/ptodata/1/ina/6B COMB.seq:    |
| 5: | /cgn2_6/ptodata/1/ina/H COMB.seq:     |
| 6: | /cgn2_6/ptodata/1/ina/PGTUS COMB.seq: |
| 7: | /cgn2_6/ptodata/1/ina/PP COMB.seq:    |
| 8: | /cgn2_6/ptodata/1/ina/RE COMB.seq:    |
| 9: | /cgn2_6/ptodata/1/ina/backfiles1.seq: |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 195.5 | 41.7        | 2029   | 3  | US-09-136-574A-46 |
| 2          | 194.5 | 41.5        | 6416   | 3  | US-09-136-574A-2  |
| 3          | 192.5 | 41.0        | 11707  | 3  | US-09-136-574A-1  |
| 4          | 182.5 | 38.9        | 2977   | 2  | US-07-862-588B-1  |
| 5          | 165   | 35.2        | 1438   | 3  | US-09-339-159B-3  |
| 6          | 165   | 35.2        | 1482   | 3  | US-09-198-956-9   |
| 7          | 165   | 35.2        | 1482   | 3  | US-09-198-955A-11 |
| 8          | 165   | 35.2        | 1482   | 3  | US-09-694-531-11  |
| 9          | 165   | 35.2        | 1482   | 3  | US-09-670-141-9   |

|    |       |      |        |   |                     |                   |
|----|-------|------|--------|---|---------------------|-------------------|
| 10 | 165   | 35.2 | 1482   | 3 | US-10-072-152-11    | Sequence 11, Appl |
| 11 | 155.5 | 33.2 | 4059   | 3 | US-09-784-554B-1    | Sequence 1, Appl  |
| 12 | 150.5 | 32.1 | 4056   | 3 | US-09-784-554B-3    | Sequence 3, Appl  |
| 13 | 120.5 | 25.7 | 1624   | 2 | US-07-862-588B-6    | Sequence 6, Appl  |
| 14 | 120.5 | 25.7 | 1775   | 2 | US-07-862-588B-5    | Sequence 5, Appl  |
| 15 | 114   | 24.3 | 486    | 2 | US-08-048-164A-1    | Sequence 1, Appl  |
| 16 | 114   | 24.3 | 486    | 2 | US-08-048-164A-3    | Sequence 3, Appl  |
| 17 | 114   | 24.3 | 486    | 2 | US-08-460-462-1     | Sequence 1, Appl  |
| 18 | 114   | 24.3 | 486    | 2 | US-08-460-462-3     | Sequence 3, Appl  |
| 19 | 114   | 24.3 | 486    | 2 | US-08-460-457-1     | Sequence 1, Appl  |
| 20 | 114   | 24.3 | 486    | 2 | US-08-460-457-3     | Sequence 3, Appl  |
| 21 | 114   | 24.3 | 486    | 2 | US-08-460-458-1     | Sequence 1, Appl  |
| 22 | 114   | 24.3 | 486    | 2 | US-08-460-458-3     | Sequence 3, Appl  |
| 23 | 114   | 24.3 | 486    | 2 | US-08-460-455-1     | Sequence 1, Appl  |
| 24 | 114   | 24.3 | 486    | 2 | US-08-460-455-3     | Sequence 3, Appl  |
| 25 | 114   | 24.3 | 486    | 2 | US-08-330-394A-1    | Sequence 1, Appl  |
| 26 | 114   | 24.3 | 486    | 2 | US-08-330-394A-3    | Sequence 3, Appl  |
| 27 | 114   | 24.3 | 499    | 3 | US-09-006-635-5     | Sequence 5, Appl  |
| 28 | 114   | 24.3 | 499    | 3 | US-09-006-632-5     | Sequence 5, Appl  |
| 29 | 114   | 24.3 | 499    | 3 | US-09-325-274-5     | Sequence 5, Appl  |
| 30 | 113   | 24.1 | 1146   | 3 | US-09-277-716-21    | Sequence 21, Appl |
| 31 | 113   | 24.1 | 1146   | 3 | US-09-609-161B-21   | Sequence 21, Appl |
| 32 | 78.5  | 16.7 | 764    | 3 | US-09-533-559-5404  | Sequence 5404, Ap |
| 33 | 77    | 16.4 | 8107   | 3 | US-09-335-586-3     | Sequence 3, Appl  |
| 34 | 76    | 16.2 | 136058 | 3 | US-09-949-016-12565 | Sequence 12565, A |
| 35 | 76    | 16.2 | 136480 | 3 | US-09-949-016-17064 | Sequence 17064, A |
| 36 | 75.5  | 16.1 | 5714   | 3 | US-09-620-312D-393  | Sequence 393, App |
| 37 | 72.5  | 15.5 | 819    | 3 | US-09-902-540-3752  | Sequence 3752, Ap |
| 38 | 72.5  | 15.5 | 17897  | 3 | US-09-902-540-1182  | Sequence 1182, Ap |
| 39 | 72.5  | 15.5 | 20284  | 3 | US-09-526-193A-21   | Sequence 21, Appl |
| 40 | 70    | 14.9 | 10892  | 3 | US-09-902-540-962   | Sequence 962, App |
| 41 | 70    | 14.9 | 25165  | 3 | US-09-453-702B-39   | Sequence 39, Appl |
| 42 | 70    | 14.9 | 25165  | 3 | US-10-114-170-39    | Sequence 39, Appl |
| 43 | 70    | 14.9 | 246444 | 3 | US-09-949-016-13113 | Sequence 13113, A |
| 44 | 69.5  | 14.8 | 103987 | 3 | US-09-949-016-12513 | Sequence 12513, A |
| 45 | 69.5  | 14.8 | 103988 | 3 | US-09-949-016-17050 | Sequence 17050, A |

#### ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

;; FILING DATE: September 19, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bak, Mary E.  
;; REGISTRATION NUMBER: 31,215  
;; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-540-9200  
;; TELEFAX: 215-540-5818  
;; TELEX: <Unknown>  
;; INFORMATION FOR SEQ ID NO: 46:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2029 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
US-09-136-574A-46

Alignment Scores:  
Pred. NO.: 5,09e-17 Length: 2029  
Score: 195.50 Matches: 38  
Percent Similarity: 60.2% Conservative: 15  
Best Local Similarity: 43.2% Mismatches: 34  
Query Match: 41.7% Indels: 1  
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-46 (1-2029)

Qy 1 ValSerGlyValValGlnTyrValAsnAspSerAlaProGlyAspAsnGln 20  
Db 1 ATGGGAAGTGTGTGAAGGTACTCTACAGAACAAATGACAGACAGTGGCAGCAGGTTCT 60  
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 61 ATAAGCGCGTGTAGATAGTAGTGAATGGAGGACAGCAGCAGTGTGATCTTAGCAGGTT 120  
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 121 AAGATAAGACTACTGGTACACAGTGGATGGTCAACAGCCACAGAGTGGCGTA--TGTGAC 177  
Qy 61 TrpAlaAlaMetGlyCysGlyValAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80  
Db 178 TGGGCACAGATAGGGGCAAGCAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTG 237  
Qy 81 ProThrAlaAspThrTyrLeuGln 88  
Db 238 AGTGGAGCGGATATTACCTGGAG 261

RESULT 2

US-09-136-574A-2  
; Sequence 2, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible  
;; OPERATING SYSTEM: DOS  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/136,574A  
;; FILING DATE: 19-Aug-1998  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/932,571  
;; FILING DATE: September 19, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bak, Mary E.  
;; REGISTRATION NUMBER: 31,215  
;; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 215-540-9200  
;; TELEFAX: 215-540-5818  
;; TELEX: <Unknown>  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 6416 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-136-574A-2

Alignment Scores:  
Pred. NO.: 3,72e-16 Length: 6416  
Score: 194.50 Matches: 38  
Percent Similarity: 61.2% Conservative: 14  
Best Local Similarity: 44.7% Mismatches: 32  
Query Match: 41.5% Indels: 1  
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-2 (1-6416)

Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 2665 GGTGTGAAGGTACTCTACAGAACAAATGACAGACAGTGGCAGCAGGTTCTATAAGCGG 2724  
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValAlaArg 43  
Db 2725 TGGTTTAAAGATAGTGAATGGAGGAGCAGCAGCAGTGTGATCTTAGCAGGGTTAAGATAAGA 2784  
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 2785 TACTGGTACACAGTGGATGGTGACAAAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 2841  
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 2842 ATAGGGCAACAAATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGAGTGGAGCG 2901  
Qy 84 AspThrTyrLeuGln 88  
Db 2902 GATTATTACCTGGAG 2916

RESULT 3

US-09-136-574A-1

; Sequence 1, Application US/09136574A

; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for

Treating Cellulose Containing Fabrics Using Truncated

Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

```

;
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11707 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-136-574A-1

Alignment Scores:
Pred. No.: 1.69e-15 Length: 11707
Score: 192.50 Matches: 37
Percent Similarity: 62.4% Conservative: 16
Best Local Similarity: 43.5% Mismatches: 31
Query Match: 41.0% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-136-574A-1 (1-11707)
QY 4 GlyValValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 4038 GGTITGAAGGTACTATACAAAGCAATGAGCAAGTGGCAGCAGCAAGTCTTATAAGGCG 4097

QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 4098 TGGTTAAGATAGTAAGTGGAGGCGAGCAGCAGTGTGATCTAGCAGGTTAAGATAAGA 4157

QY 44 TyrIrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 4158 TACTGGTACACAGTGGATGGTGACAAAGCCACAGAGTGGCGTA--TGTGACTGGGCACAG 4214

QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 4215 ATAGGGCAACAAATGTACATTCATTTTGTGAAGCTGAGCAGCGGAGTGAGTGGAGCG 4274

QY 84 AspThrTyrLeuGln 88
Db 4275 GATTATTACTTGGAG 4289

RESULT 4
US-07-862-588B-1
; Sequence 1, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
;
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2977 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 677..2776
; OTHER INFORMATION:
;
; US-07-862-588B-1
Alignment Scores:
Pred. No.: 5.76e-15 Length: 2977
Score: 182.50 Matches: 36
Percent Similarity: 66.3% Conservative: 19
Best Local Similarity: 43.4% Mismatches: 25
Query Match: 38.9% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-1 (1-2977)
QY 6 LysValGlnTyrIleAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25
Db 144 CGACTGCAGTACAGCGCGCGCATACAAATGACGAGCGGACCAACCATGATCAAGCGCTCCTTC 203

QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45
Db 204 AACATCAAAACAAACACGGTACTTCGGCTGTGATTAAAGCAGCTCAAAATCCGCTACTAC 263

QY 46 PheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65
Db 264 TTCACCAAGATGGTCTCGCGCGGTGAACGGTGG---ATCGACTGGCGCGCAGCTCGGC 320

QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85
Db 321 GCAGCAACATTCAGATCTCGTTTGGC-----AACCATACTGCGCAGCAATTCGATACG 374

QY 86 TyrLeuGln 88
```



```
|||||
Db 375 TACGTGGAG 383

RESULT 5
US-09-339-159B-3
; Sequence 3, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; PRIORITY FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-09-339-159B-3

Alignment Scores:
Pred. No.: 5,53e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-4 (1-89) x US-09-339-159B-3 (1-1438)
Qy 1 ValSerGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGTTGAATTTCTACACAGCAATCTTCAGATACTACTACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAAGTTACTTAATACCGAAGCAGTGCATTTGTCTCAAACTC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGCGAGCAAGATCAGACCTTCTGCG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTAATCTCAAAATGTAAGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGAGACACCTTACCTTGAA 1227

RESULT 6
US-09-198-956-9
; Sequence 9, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; PRIORITY FILING DATE: 1998-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240

US-09-917-376-4 (1-89) x US-09-198-956-9 (1-1482)
Qy 1 ValSerGlyValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTTGAAGTTGAATTTCTACACAGCAATCTTCAGATACTACTACTCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAAGTTACTTAATACCGAAGCAGTGCATTTGTCTCAAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGCGAGCAAGATCAGACCTTCTGCG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGAGCTACACGGAATTAATCTCAAAATGTAAGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTTGTAATAATGAGTTCTCTCAACAAATAACGAGACACCTTACCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; PRIORITY FILING DATE: 1998-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,249
; PRIORITY FILING DATE: 1997-12-02
; EARLIER FILING DATE: 1997-12-02
; EARLIER FILING DATE: 1997-12-02
; EARLIER FILING DATE: 1998-05-06
; EARLIER FILING DATE: 1998-05-06
; EARLIER FILING DATE: 1998-11-02
; EARLIER FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
```

| Qy | 1   | Val  | Ser | Gly  | Gly  | Val | Lys  | Val  | Leu  | Tyr | Lys  | Asn  | Asp | Ser  | Ala  | Pro  | Gly  | Asp  | Asn  | Gln  | 20   |
|----|-----|------|-----|------|------|-----|------|------|------|-----|------|------|-----|------|------|------|------|------|------|------|------|
| Db | 991 | GTAT | CAG | CAAT | TGTA | AGG | TTGA | ATTT | CTAC | AAG | CAAT | CTCT | CAG | ATCT | CTCT | AACT | CTCT | CTCT | CTCT | CTCT | 1050 |

```
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTTCAAGGTTACTAATACCGAAGCAGTGCATTTGTTGTCAAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAAATCGGCAGTAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278
RESULT 10
US-09-784-152-11
; Sequence 11, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11
Alignment Scores:
Pred. No.: 5.78e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 3 Gaps: 2
US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCGCAATTGAAGGTTGAATTTCAACAGCAATCTTCCAGATACCTACTCACTCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTTCAAGGTTACTAATACCGAAGCAGTGCATTTGTTGTCAAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAAATCGGCAGTAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278
RESULT 10
US-09-784-152-11
; Sequence 11, Application US/10072152
; Patent No. 6677147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6677147el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-072-152-11
Alignment Scores:
Pred. No.: 5.78e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 3 Gaps: 2
US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)
Qy 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCGCAATTGAAGGTTGAATTTCAACAGCAATCTTCCAGATACCTACTCACTCA 1050
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTTCAAGGTTACTAATACCGAAGCAGTGCATTTGTTGTCAAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAAATCGGCAGTAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278
RESULT 11
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1
Alignment Scores:
Pred. No.: 5.25e-11 Length: 4059
Score: 155.50 Matches: 27
Percent Similarity: 56.5% Conservative: 21
Best Local Similarity: 31.8% Mismatches: 36
Query Match: 33.2% Indels: 1
DB: 3 Gaps: 1
US-09-917-376-4 (1-89) x US-09-784-554B-1 (1-4059)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGAATGCTGCTCAATATCGCACAGCAGTAACTAGGTGAACAGATAATCACCTCAATCCG 3657
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3658 CAATTCCAAATTGTAAACAAGGACACCTCCATACCGATCAACGAGTTGAAATTCGC 3717
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3718 TACTACTACACAATCGACGGTGCACCGTGAGCAG---ACTTTCACACTGCGACTATGCGACG 3774
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAGCTGAACGGTTAACTGGTTAAATGAGAGGCTGCAACCGGTGCC 3834
Qy 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849
RESULT 12
US-09-784-554B-3
; Sequence 3, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
```

```
Db 1051 ATCAATCCTCAGTTTCAAGGTTACTAATACCGAAGCAGTGCATTTGTTGTCAAACTC 1110
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTACGCGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAAATCGGCAGTAACGCGAGCTACAACGGAATTACTTCAAAATGTAAAGGA 1227
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTGAAAAATGAGTTCTCTCAACAAATAACGCGACACCTTACCTTGAA 1278
RESULT 11
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1
Alignment Scores:
Pred. No.: 5.25e-11 Length: 4059
Score: 155.50 Matches: 27
Percent Similarity: 56.5% Conservative: 21
Best Local Similarity: 31.8% Mismatches: 36
Query Match: 33.2% Indels: 1
DB: 3 Gaps: 1
US-09-917-376-4 (1-89) x US-09-784-554B-1 (1-4059)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3598 GGAATGCTGCTCAATATCGCACAGCAGTAACTAGGTGAACAGATAATCACCTCAATCCG 3657
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3658 CAATTCCAAATTGTAAACAAGGACACCTCCATACCGATCAACGAGTTGAAATTCGC 3717
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 3718 TACTACTACACAATCGACGGTGCACCGTGAGCAG---ACTTTCACACTGCGACTATGCGACG 3774
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3775 CTGAGCTGCTCAAGCTGAACGGTTAACTGGTTAAATGAGAGGCTGCAACCGGTGCC 3834
Qy 84 AspThrTyrLeuGln 88
Db 3835 GATTATTATTGGAA 3849
RESULT 12
US-09-784-554B-3
; Sequence 3, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
```

```
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784.554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 2.61e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.5% Conservative: 21
Best Local Similarity: 31.8% Mismatches: 36
Query Match: 32.1% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-4 (1-89) x US-09-784-554B-3 (1-4056)
QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 3595 CGATTGGTCTCCAGTATCGCACGGGATCAAAATGTGAACGACATCATTGACCCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 3655 CATTTCCAAATTTAAATAAAGGTACATCTCCGTACCGATCAACGAGTTGAAATTCGC 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAspTrpAlaAla 63
DB 3715 TACTACTACAGCATCGACGGTGACCGTGAGCAG---ACATTCAACTGCGACTATCGCGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 3772 CTGAGCTGCTCGAAGCTCAATGTAAGTGTAAATGGATAAGCTGCAACCGTGCT 3831
QY 84 AspThrTyrLeuGln 88
DB 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991

; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1.06e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.1% Conservative: 18
Best Local Similarity: 32.9% Mismatches: 36
Query Match: 25.7% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-4 (1-89) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGAACTTGTGTTCCTCAATCAAAAGTTGGGACACTAGCCGCGGATAACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AAGCTTCTTTAAACATCAAGAACACCGGTACAAACCCCTGTTAACTGACGGCGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAsp 61
DB 1443 CTNNNNNNNNNNNNNNNAAAAGAC---GGACCTGGCGGATATGAGCTGCTCGATCGACTGG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCGCAAAATCGGCGGAAAGAAATGTTCTGCTGGCATTG-----GCTAACTTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGGATACCTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linea
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..(1625..1775)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
US-07-862-588B-5

```

```

Alignment Scores:
Pred. No.: 1.2e-06 Length: 1775
Score: 120.50 Matches: 28
Percent Similarity: 54.1% Conservative: 18
Best Local Similarity: 32.9% Mismatches: 36
Query Match: 25.7% Indels: 3
DB: 2 Gaps: 2

```

```

US-09-917-376-4 (1-89) x US-07-862-588B-5 (1-1775)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 1323 ACGGGGAACCTGTGTTCACAAACAAAGTTGGCGACACTACGCGCACCGGATCAACAAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 1383 AAGCCTTCCTTTACATCAAGACACGATACACCCCTGTACCTGAGCGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
DB 1443 CTNNNNNNNNNNNNNNAAAAAGAC---GGACCTGGGATATGAGTCTCGATCGACTGG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
DB 1500 GCCCAATCGCGCAACGAATGTTCTCTGGCATTC-----GCTAACTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
DB 1554 AATACGGATACTTAC 1568
RESULT 15
US-08-048-164A-1

```

```

; Sequence 1, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shpiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-048-164A-1
Alignment Scores:
Pred. No.: 1.49e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.6% Conservative: 22
Best Local Similarity: 29.9% Mismatches: 34
Query Match: 24.3% Indels: 12
DB: 2 Gaps: 4
US-09-917-376-4 (1-89) x US-08-048-164A-1 (1-486)
QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
DB 7 ACATCATCAATGTCAGTTGAATTTTACAACTCTTACAAATCAGCAACAACTCAAT 66
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
DB 67 ACACCAATATCAAAATTTACTAACACATCTCACAGTGATTTAAATTTAAATGACGTA 126
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
DB 127 GTTAGATATTATTACACAAGTGATGTGTACACAAGCAAACTTTCTGG---TGTGACCAT 183
QY 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
DB 184 GCTGTCATATTATTAGGAAATAGCTGTGTGATACTAGCAAAAGTACACAACTTC 243
QY 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88

```

Db 244 GTTAAAGAAACAGCAAGCCCA--ACATCAACCTATGATACATATGTTGAA 291

Search completed: March 2, 2006, 13:50:30  
Job time : 38.1046 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:19:26 ; Search time 167.029 Seconds  
(without alignments)  
4406.259 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 469

Sequence: 1 VSGGVKQVKNDSAPGDNQ.....RASFGSVNPATPTADTYLQX 89

Scoring table:

|                           |  |
|---------------------------|--|
| BLOSUM62                  |  |
| Xgapop 10.0 , Xgapext 0.5 |  |
| Ygapop 10.0 , Ygapext 0.5 |  |
| Fgapop 6.0 , Fgapext 7.0  |  |
| Delop 6.0 , Delext 7.0    |  |

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/abs/ABSSWEB/spool/US09917376/runat\_02032006\_091500\_8223/app\_query.fasta\_1  
-DB=Published\_Applications\_NA\_Main -QFMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=BLOSUM62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs06p  
-USER=US09917376 @CGN\_1\_1 3095 @runat\_02032006\_091500\_8223 -NCPUP=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA\_Main:

|     |  |
|-----|--|
| 1:  | /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*  |
| 2:  | /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*  |
| 3:  | /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.* |
| 4:  | /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.* |
| 5:  | /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.* |
| 6:  | /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.* |
| 7:  | /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.* |
| 8:  | /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.* |
| 9:  | /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.* |
| 10: | /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Query Match | Score | Length  | ID | Description       |
|------------|-------------|-------|---------|----|-------------------|
| 1          | 469         | 100.0 | 2289    | 3  | US-09-917-378-2   |
| 2          | 466         | 99.4  | 3687    | 3  | US-09-917-384-2   |
| 3          | 466         | 99.4  | 3687    | 3  | US-09-917-383-2   |
| 4          | 462         | 98.5  | 2869    | 3  | US-09-917-376-2   |
| 5          | 462         | 98.5  | 2869    | 5  | US-10-155-400-2   |
| 6          | 257         | 54.8  | 2223    | 6  | US-10-156-761-550 |
| 7          | 257         | 54.8  | 9025608 | 6  | US-10-156-761-1   |

|    |       |      |       |   |                      |
|----|-------|------|-------|---|----------------------|
| 8  | 190   | 40.5 | 1527  | 6 | US-10-369-493-46838  |
| 9  | 188   | 40.1 | 1434  | 8 | US-10-466-208-11     |
| 10 | 188   | 40.1 | 1491  | 8 | US-10-466-208-7      |
| 11 | 188   | 40.1 | 2510  | 7 | US-10-433-577-34     |
| 12 | 165   | 35.2 | 501   | 9 | US-10-933-404-1      |
| 13 | 165   | 35.2 | 1438  | 6 | US-10-372-054-3      |
| 14 | 165   | 35.2 | 1482  | 5 | US-10-072-152-11     |
| 15 | 165   | 35.2 | 1482  | 7 | US-10-655-433-11     |
| 16 | 157   | 33.5 | 1314  | 3 | US-09-974-300-684    |
| 17 | 155.5 | 33.2 | 4059  | 3 | US-09-784-554B-1     |
| 18 | 155.5 | 33.2 | 4059  | 3 | US-10-896-555-1      |
| 19 | 150.5 | 32.1 | 4056  | 3 | US-09-784-554B-3     |
| 20 | 150.5 | 32.1 | 4056  | 8 | US-10-896-555-3      |
| 21 | 131.5 | 28.0 | 3180  | 9 | US-10-886-393A-29    |
| 22 | 130.5 | 27.8 | 4452  | 7 | US-10-282-122A-15299 |
| 23 | 126.5 | 27.0 | 1303  | 9 | US-10-490-737-1      |
| 24 | 126.5 | 27.0 | 1747  | 9 | US-10-490-737-2      |
| 25 | 126.5 | 27.0 | 3489  | 9 | US-10-490-737-3      |
| 26 | 114   | 24.3 | 768   | 6 | US-10-261-446-5      |
| 27 | 114   | 24.3 | 768   | 7 | US-10-261-445B-5     |
| 28 | 114   | 24.3 | 768   | 8 | US-10-782-234-5      |
| 29 | 113   | 24.1 | 652   | 8 | US-10-748-055-9      |
| 30 | 113   | 24.1 | 1146  | 3 | US-09-808-898-21     |
| 31 | 78.5  | 16.7 | 764   | 8 | US-10-653-047-5404   |
| 32 | 78.5  | 16.7 | 3147  | 7 | US-10-282-122A-13716 |
| 33 | 76    | 16.2 | 757   | 7 | US-10-767-701-8547   |
| 34 | 76    | 16.2 | 96597 | 7 | US-10-052-482-226    |
| 35 | 75.5  | 16.1 | 1041  | 7 | US-10-260-238-742    |
| 36 | 75.5  | 16.1 | 1041  | 7 | US-10-437-963-76751  |
| 37 | 75.5  | 16.1 | 2091  | 7 | US-10-211-462-222    |
| 38 | 75.5  | 16.1 | 2091  | 8 | US-10-723-860-4038   |
| 39 | 75.5  | 16.1 | 2091  | 9 | US-10-756-149-4006   |
| 40 | 75.5  | 16.1 | 3396  | 7 | US-10-398-458-1      |
| 41 | 75.5  | 16.1 | 4041  | 3 | US-09-747-835A-21    |
| 42 | 75.5  | 16.1 | 4041  | 7 | US-10-312-312-21     |
| 43 | 75.5  | 16.1 | 4353  | 5 | US-10-225-567A-405   |
| 44 | 75.5  | 16.1 | 4755  | 9 | US-10-505-486-201    |
| 45 | 75.5  | 16.1 | 4914  | 3 | US-09-974-298-188    |

#### ALIGNMENTS

##### RESULT 1

US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication No. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS  
; FILE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

|                        |          |               |      |
|------------------------|----------|---------------|------|
| Alignment Scores:      | 6.94e-56 | Length:       | 2289 |
| Pred. No.:             | 469.00   | Matches:      | 88   |
| Score:                 | 100.0%   | Conservative: | 0    |
| Best Local Similarity: | 100.0%   | Mismatches:   | 0    |
| Query Match:           | 100.0%   | Indels:       | 0    |
| DB:                    | 3        | Gaps:         | 0    |





```

; CURRENT APPLICATION NUMBER: US/09/917,376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-09-917-376-2

Alignment Scores:
Pred. No.: 9,34e-55 Length: 2869
Score: 462.00 Matches: 86
Percent Similarity: 100.0% Conservative: 2
Best Local Similarity: 97.7% Mismatches: 0
Query Match: 98.5% Indels: 0
DB: 3 Gaps: 0

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGAAGGTGCAGTATAAGATAATGATTCGGCGCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTCGAGGTGGTGAATACCGGGTCGTGCGTGGATTGTCCACCGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGGGTACTGTTTACCCGGGATGGTGGCTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGGGCGCGATCGGGTGTGGGAATATCCGGCCTCGTTCGGTCCGTTGAAACCCGCGCAG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACTACCTGCAG 2868

RESULT 5
US-10-155-400-2
; Sequence 2, Application US/10155400
; Publication No. US2003010898A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2869
; TYPE: DNA
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2869)
; OTHER INFORMATION: a, c, t, g, other or unknown
US-10-155-400-2

Alignment Scores:
Pred. No.: 9,34e-55 Length: 2869
Score: 462.00 Matches: 86

US-09-917-376-4 (1-89) x US-09-917-376-2 (1-2869)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 2605 GTGTCGGGTGGGGTGAAGGTGCAGTATAAGATAATGATTCGGCGCGGGTGATAATCAG 2664

QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2665 ATCAAGCCGGGTTCGAGGTGGTGAATACCGGGTCGTGCGTGGATTGTCCACCGTG 2724

QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 2725 ACGGTGGGTACTGTTTACCCGGGATGGTGGCTCGACACTGGTGACACTGTGAC 2784

QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 2785 TGGGGCGCGATCGGGTGTGGGAATATCCGGCCTCGTTCGGTCCGTTGAAACCCGCGCAG 2844

QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 2845 CCGACGGCGGACACTACCTGCAG 2868

RESULT 6
US-10-156-761-550
; Sequence 550, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 550
; LENGTH: 2223
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2223)
US-10-156-761-550

Alignment Scores:
Pred. No.: 1.17e-25 Length: 2223
Score: 257.00 Matches: 49
Percent Similarity: 71.3% Conservative: 13
Best Local Similarity: 56.3% Mismatches: 25
Query Match: 54.8% Indels: 0
DB: 6 Gaps: 0

US-09-917-376-4 (1-89) x US-10-156-761-550 (1-2223)

QY 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 21
Db 1771 TCGGGCGGCTCAAGGTCTCTACAGAGAACAACTCTCTCGGCCCGCCGCAACGCGCATC 1830

QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 2223 TCGGGCGGCTCAAGGTCTCTACAGAGAACAACTCTCTCGGCCCGCCGCAACGCGCATC 1830
```

Db 1831 CGGCCAGCCTTCGATCGTCAACACCGCGCAGCGCTCCCTCGACCTGTCACAGGTACG 1890  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 1891 GCCCGCTACTACTTACCCGGGACACCGCGCTCGCCACCGCTGAGCGCTGCGACTAC 1950  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 1951 CGGCCGCTGCGCTGTTCCAAAGTCAGCTGAGGTGCTGACCCCTGACCACCGCCCGTCCG 2010  
QY 82 ThrAlaAspThrTyrLeuGln 88  
Db 2011 GGAGCGGAGCTACTCTCGAA 2031  
RESULT 7  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1  
Alignment Scores:  
Pred. No.: 7,4e-21 Length: 9025608  
Score: 257.00 Matches: 49  
Percent Similarity: 71.3% Conservative: 13  
Best Local Similarity: 56.3% Mismatches: 25  
Query Match: 54.8% Indels: 0  
DB: 6 Gaps: 0  
US-09-917-376-4 (1-89) x US-10-156-761-1 (1-9025608)  
QY 2 SerGlyGlyValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIle 21  
Db 706216 TCGGGCGCCTCAGGTCTCTACAGAACACGACTCTCGGCCACCGACCGCCATC 706275  
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41  
Db 706276 CGGCCAGCCTTCGATCGTCAACACCGCGCGCTCCCTCGACCTGTCCAAGGTACG 706335  
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61  
Db 706336 GCCCGCTACTACTTACCCGGGACACCGCGCTCGCCACCGCTGAGCGCTGCGACTAC 706395  
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
Db 706396 CGGCCGCTGCGCTGTTCCAAAGTCAGCTGAGGTGCTGACCCCTGACCACCGCCCGTCCG 706455  
QY 82 ThrAlaAspThrTyrLeuGln 88

Db 706456 GGAGCGGAGCCTACTCTGAA 706476  
RESULT 8  
US-10-369-493-46838  
; Sequence 46838, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 46838  
; LENGTH: 1527  
; TYPE: DNA  
; ORGANISM: Bacillus subtilis  
US-10-369-493-46838  
Alignment Scores:  
Pred. No.: 2,55e-16 Length: 1527  
Score: 190.00 Matches: 35  
Percent Similarity: 65.1% Conservative: 21  
Best Local Similarity: 40.7% Mismatches: 26  
Query Match: 40.5% Indels: 4  
DB: 6 Gaps: 2  
US-09-917-376-4 (1-89) x US-10-369-493-46838 (1-1527)  
QY 4 GlyValLysValGlnTyrIleAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
Db 1084 GGTATTCTGTACAGTACAGCAGCGGAGTGGAGTATGACAGCAACCAATCCGTCG 1143  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerSerValAspLeuSerThrValThrValArg 43  
Db 1144 CAGCTTCAATAAAAAAATACCGCAATACCGGTTGATTAAAAAGATGCTACTGCCGT 1203  
QY 44 TyrTrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAla 62  
Db 1204 TACTGTATAAGCGAAACAAAGGCCAAAC-----TTTACTGTGACTACGCG 1254  
QY 63 AlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThr 82  
Db 1255 CAGATTGGATCGGCAATGTGACACACAAAGTTTGTGACGTTCATATAACCAAGCAAGT 1314  
QY 83 AlaAspThrTyrLeuGln 88  
Db 1315 GCAGATACCTATCTCGAA 1332  
RESULT 9  
US-10-466-208-11  
; Sequence 11, Application US/10466208  
; Publication No. US20040180348A1  
; GENERAL INFORMATION:  
; APPLICANT: GENOFOCUS Co., Ltd.  
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers  
; FILE REFERENCE: 3260-13  
; CURRENT APPLICATION NUMBER: US/10/466,208  
; CURRENT FILING DATE: 2003-07-14  
; PRIOR APPLICATION NUMBER: KR2001-2156  
; PRIOR FILING DATE: 2001-01-15  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Kopatentin 1.71  
; SEQ ID NO 11  
; LENGTH: 1434  
; TYPE: DNA

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCcase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11
Alignment Scores:
Pred. No.: 4.53e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 8 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-11 (1-1434)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTCTGTACAATACAGACAGGAGTGGAGTATGAACAGCAACCAATCCGTCG 1050
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATAAATAAATAACGCAATACCAACCGTTGATTAAAGATGTCACCTGCCCGT 1110
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 1111 TACTGTATTAACGCGAATAAAGCCAAACAGTT-----GACTGTGACTACGCGCAG 1164
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTGGATCGGCAATGTGACATACAAGTTTGTGACGTTCGATATAACCAACCAAGAGGTGCA 1224
Qy 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239
RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180349A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCcase gene with mutated signal sequence to enhance its
; OTHER INFORMATION: hydrophobicity
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7
Alignment Scores:
Pred. No.: 4.77e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 8 Gaps: 1
US-09-917-376-4 (1-89) x US-10-466-208-7 (1-1491)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 991 GGTATTCTGTACAATACAGACAGGAGTGGAGTATGAACAGCAACCAATCCGTCG 1050
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 1051 CAGCTTCAATAAATAAATAACGCAATACCAACCGTTGATTAAAGATGTCACCTGCCCGT 1110
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 1111 TACTGTATTAACGCGAATAAAGCCAAACAGTT-----GACTGTGACTACGCGCAG 1164
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1165 CTGGATCGGCAATGTGACATACAAGTTTGTGACGTTCGATATAACCAACCAAGAGGTGCA 1224
Qy 84 AspThrTyrLeuGln 88
Db 1225 GATACCTATCTGGAA 1239
RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34
Alignment Scores:
Pred. No.: 9.54e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 7 Gaps: 1
US-09-917-376-4 (1-89) x US-10-433-577-34 (1-2510)
Qy 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2051 GGTATTCTGTACAATACAGACAGGAGTGGAGTATGAACAGCAACCAATCCGTCG 2110
Qy 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2111 CAGCTTCAATAAATAAATAACGCAATACCAACCGTTGATTAAAGATGTCACCTGCCCGT 2170
Qy 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 2171 TACTGTATTAACGCGAATAAAGCCAAACAGTT-----GACTGTGACTACGCGCAG 2224
Qy 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2225 CTGGATCGGCAATGTGACATACAAGTTTGTGACGTTCGATATAACCAACCAAGAGGTGCA 2284
Qy 84 AspThrTyrLeuGln 88
Db 2285 GATACCTATCTGGAA 2299
```

```

; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. I633
US-10-372-054-3

Alignment Scores:
Pred. No.:      8.68e-13      Length:      1438
Score:          165.00      Matches:      35
Percent Similarity: 59.8%      Conservative: 23
Best Local Similarity: 36.1%      Mismatches:  29
Query Match:     35.2%      Indels:      10
DB:              6          Gaps:       2

US-09-917-376-4 (1-89) x US-10-372-054-3 (1-1438)

Qy 1 ValSerGlyClyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCATGCCAATTTGAAGGTTGAATTTCTACACAGCAATCCTTCAGATACTACTCA 999
Qy 21 ILsLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAAGGTTACTAATACCGGAAGCAGTGCATTTGATTGTGCCAAATC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATATTATTATACGTAGTACGGACAGAAAGATCAGACCTTCTGCG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGTTCGAATAATCGGCAGTAAACGGCAGCTACACGGAAATTAATCTCAATGTAAAGGA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTTGTAAAAATGATGTTCTCTCAACAAATAACCCGACACACCTTACCTTGAA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schlein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum

```

## US-10-072-152-11

## Alignment Scores:

Pred. No.: 9,048-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservative: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 5 Gaps: 2

US-09-917-376-4 (1-89) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 991 GTATCAGGCAATTGAAGTTGAATCTACACAGCAATCCTTCAGATACTACTAATCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATGATTGTCCAAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1111 ACATTGAGATATTATTATACAGTACGACGACGAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
DB 1168 CATGTCGAATAATCGGCAGTAAACGCGAGTACACAGCAATTAATCTCAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
DB 1228 ACATTGTAAAATGAGTTCTTCAACAAATAATACGCGACACCTACCTTGAA 1278

## RESULT 15

US-10-655-433-11  
; Sequence 11, Application US/10655433  
; Publication No. US20040067572A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Made E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnort, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/10/655,433  
; CURRENT FILING DATE: 2003-09-04  
; PRIOR APPLICATION NUMBER: US/09/198,955A  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-10-655-433-11

## Alignment Scores:

Pred. No.: 9,048-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservative: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 7 Gaps: 2

US-09-917-376-4 (1-89) x US-10-655-433-11 (1-1482)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
DB 991 GTATCAGGCAATTGAAGTTGAATCTACACAGCAATCCTTCAGATACTACTAATCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
DB 1051 ATCAATCTCAGTTCAAGTTACTAATACCGAAGCAGTGCAATGATTGTCCAAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
DB 1111 ACATTGAGATATTATTATACAGTACGACGACGAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
DB 1168 CATGTCGAATAATCGGCAGTAAACGCGAGTACACAGCAATTAATCTCAATGTAAAGGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
DB 1228 ACATTGTAAAATGAGTTCTTCAACAAATAATACGCGACACCTACCTTGAA 1278

Search completed: March 2, 2006, 20:17:40  
Job time : 697.029 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:24:59 ; Search time 95.5064 Seconds  
(without alignments)  
2043.020 Million cell updates/sec

Title: US-09-917-376-4

Perfect score: 459

Sequence: 1 VSGGVKQVKNDSAPGDNO.....RASFGSVNRPATPDYVLQX 89

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7218512 seqs, 1096188070 residues

Total number of hits satisfying chosen parameters: 14437024

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09917376/runat\_02032006\_091502\_8293/app\_query.fasta.1  
-DB=Published Applications NA\_New -QFWT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs03p  
-USER=US09917376 -CGN\_1\_1\_1067@runat\_02032006\_091502\_8293 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA\_New:

|  | 1:  | /cgn2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq.*  |
|--|-----|--|
|  | 2:  | /cgn2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq.*  |
|  | 3:  | /cgn2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq.*  |
|  | 4:  | /cgn2_6/ptodata/2/pubnpa/PCT_NEW_PUB.seq.*   |
|  | 5:  | /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq.*  |
|  | 6:  | /cgn2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq.*  |
|  | 7:  | /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq.*  |
|  | 8:  | /cgn2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq.*  |
|  | 9:  | /cgn2_6/ptodata/2/pubnpa/US11_NEW_PUB.seq.*  |
|  | 10: | /cgn2_6/ptodata/2/pubnpa/US11_NEW_PUB.seq2.* |
|  | 11: | /cgn2_6/ptodata/2/pubnpa/US11_NEW_PUB.seq3.* |
|  | 12: | /cgn2_6/ptodata/2/pubnpa/US11_NEW_PUB.seq4.* |
|  | 13: | /cgn2_6/ptodata/2/pubnpa/US60_NEW_PUB.seq.*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID               | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1          | 163   | 34.8        | 1500   | 8 US-10-517-939-327 | Sequence 327, App |
| 2          | 157   | 33.5        | 2602   | 8 US-10-510-386-229 | Sequence 229, App |
| 3          | 113   | 24.1        | 1146   | 12 US-11-179-411-21 | Sequence 21, Appl |
| 4          | 113   | 24.1        | 1146   | 12 US-11-175-766-21 | Sequence 21, Appl |

|    |      |      |        |    |                       |                    |
|----|------|------|--------|----|-----------------------|--------------------|
| 5  | 108  | 23.0 | 2217   | 8  | US-10-517-939-343     | Sequence 343, App  |
| 6  | 70   | 14.9 | 15876  | 12 | US-11-052-554A-660    | Sequence 660, App  |
| 7  | 69.5 | 14.8 | 617    | 6  | US-09-925-065A-353393 | Sequence 353393, A |
| 8  | 68.5 | 14.6 | 1084   | 8  | US-10-750-623-31182   | Sequence 31182, A  |
| 9  | 68.5 | 14.6 | 1084   | 8  | US-10-750-623-31182   | Sequence 31182, A  |
| 10 | 68.5 | 14.6 | 2231   | 9  | US-11-072-512-855     | Sequence 855, App  |
| 11 | 68.5 | 14.6 | 4952   | 12 | US-11-136-527-487     | Sequence 487, App  |
| 12 | 68   | 14.5 | 2268   | 8  | US-10-517-939-329     | Sequence 329, App  |
| 13 | 68   | 14.5 | 35344  | 8  | US-10-995-561-13307   | Sequence 13307, A  |
| 14 | 67.5 | 14.4 | 556    | 6  | US-09-925-065A-228751 | Sequence 228751, A |
| 15 | 67.5 | 14.4 | 2564   | 9  | US-11-072-512-279     | Sequence 279, App  |
| 16 | 67   | 14.3 | 523    | 6  | US-09-925-065A-39541  | Sequence 39541, A  |
| 17 | 67   | 14.3 | 171486 | 12 | US-11-121-086-105     | Sequence 105, App  |
| 18 | 66.5 | 14.2 | 600    | 8  | US-10-750-185-1174    | Sequence 1174, App |
| 19 | 66.5 | 14.2 | 600    | 8  | US-10-750-623-1174    | Sequence 1174, App |
| 20 | 66.5 | 14.2 | 1847   | 12 | US-11-000-688-1099    | Sequence 1099, App |
| 21 | 66.5 | 14.2 | 2565   | 8  | US-10-750-185-48986   | Sequence 48986, A  |
| 22 | 66.5 | 14.2 | 2565   | 8  | US-10-750-623-48986   | Sequence 48986, A  |
| 23 | 65.5 | 14.0 | 617    | 6  | US-09-925-065A-353394 | Sequence 353394, A |
| 24 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547306 | Sequence 547306, A |
| 25 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547307 | Sequence 547307, A |
| 26 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547308 | Sequence 547308, A |
| 27 | 64.5 | 13.8 | 600    | 6  | US-09-925-065A-336480 | Sequence 336480, A |
| 28 | 64.5 | 13.8 | 600    | 6  | US-09-925-065A-336481 | Sequence 336481, A |
| 29 | 64.5 | 13.8 | 601    | 6  | US-09-925-065A-729386 | Sequence 729386, A |
| 30 | 64.5 | 13.8 | 2370   | 6  | US-09-925-065A-91998  | Sequence 91998, A  |
| 31 | 64.5 | 13.8 | 2370   | 6  | US-09-925-065A-91999  | Sequence 91999, A  |
| 32 | 64.5 | 13.8 | 159138 | 8  | US-10-995-561-13230   | Sequence 13230, A  |
| 33 | 64   | 13.6 | 556    | 6  | US-09-925-065A-369647 | Sequence 369647, A |
| 34 | 64   | 13.6 | 585    | 6  | US-09-925-065A-385747 | Sequence 385747, A |
| 35 | 64   | 13.6 | 611    | 6  | US-09-925-065A-658978 | Sequence 658978, A |
| 36 | 64   | 13.6 | 659    | 6  | US-09-925-065A-87693  | Sequence 87693, A  |
| 37 | 64   | 13.6 | 659    | 6  | US-09-925-065A-87694  | Sequence 87694, A  |
| 38 | 64   | 13.6 | 1898   | 9  | US-11-072-512-1678    | Sequence 1678, App |
| 39 | 64   | 13.6 | 1990   | 12 | US-11-062-225-2       | Sequence 2, Appl   |
| 40 | 64   | 13.6 | 1990   | 12 | US-11-062-225-10      | Sequence 10, Appl  |
| 41 | 64   | 13.6 | 3074   | 12 | US-11-000-688-315     | Sequence 315, App  |
| 42 | 63.5 | 13.5 | 596    | 6  | US-09-925-065A-521812 | Sequence 521812, A |
| 43 | 63.5 | 13.5 | 597    | 6  | US-09-925-065A-258810 | Sequence 258810, A |
| 44 | 63.5 | 13.5 | 597    | 6  | US-09-925-065A-258811 | Sequence 258811, A |
| 45 | 63.5 | 13.5 | 606    | 6  | US-09-925-065A-484851 | Sequence 484851, A |

#### ALIGNMENTS

RESULT 1  
US-10-517-939-327  
; Sequence 327, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Blum, David  
; APPLICANT: Esteghalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 327  
; LENGTH: 1500  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:

OTHER INFORMATION: Obtained from an environmental sample.  
US-10-517-939-327

Alignment Scores:  
Pred. No.: 8.87e-11 Length: 1500  
Score: 163.00 Matches: 30  
Percent Similarity: 60.0% Conservative: 21  
Best Local Similarity: 35.3% Mismatches: 30  
Query Match: 34.8% Indels: 4  
DB: 8 Gaps: 2

US-09-917-376-4 (1-89) x US-10-517-939-327 (1-1500)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24  
Db 1060 CTATCTGTGAATACAGACAGGGGAGTGAACACCAACCAATCCGCTCAG 1119  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
Db 1120 ATCCATGTGAATAACACAGCAAGCACCCTTAATTTAAATAATGTAACCTCGCTAC 1179  
Qy 45 TrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 1180 TGGTATACACGAAACAAAGGCCAAAC-----TTCGACTGTGACTACCGAAG 1230  
Qy 64 MetGlyCysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 1231 ATCGGATGACCAATGTGACGACCAAGTTTGTGCATTAACAAACCTGTAAGAGGTGCA 1290  
Qy 84 AspThrTyrLeuGln 88  
Db 1291 GATGCTATCTGGAA 1305

## RESULT 2

US-10-510-386-229  
Sequence 229, Application US/10510386  
Publication No. US20050244922A1  
GENERAL INFORMATION:  
APPLICANT: Andersen, Jens Tonne  
APPLICANT: Clausen, Ib Groth  
APPLICANT: Jorgensen, Steen Troels  
APPLICANT: Olsen, Peter Bjarke  
APPLICANT: Rasmussen, Michael Dolberg  
TITLE OF INVENTION: Improved Bacillus Host Cell  
FILE REFERENCE: 10294.204-US  
CURRENT APPLICATION NUMBER: US/10/510,386  
CURRENT FILING DATE: 2004-10-04  
NUMBER OF SEQ ID NOS: 248  
SOFTWARE: PatentIn version 3.3  
SEQ ID NO 229  
LENGTH: 2602  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (501)..(2102)

US-10-510-386-229

Alignment Scores:  
Pred. No.: 9.84e-10 Length: 2602  
Score: 157.00 Matches: 32  
Percent Similarity: 56.5% Conservative: 16  
Best Local Similarity: 37.6% Mismatches: 33  
Query Match: 33.5% Indels: 4  
DB: 8 Gaps: 2

US-09-917-376-4 (1-89) x US-10-510-386-229 (1-2602)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24  
Db 1665 ATGACAGTACATACAGAGCGGGGAGCAACAATGTAACGCAACCAATCCGCTCAG 1724  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44

Db 1725 CTCACATTAAACAAACAGCAAAAAACCGGTCTTTAAATCGAATCACTGTCCGCTAC 1784  
Qy 45 TrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
Db 1785 TGGTATAAACGAATCGCAAGGACAAAAT-----TTTGACTGCGACTATGCCCAA 1835  
Qy 64 MetGlyCysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
Db 1836 ATCGCTGCAGCAAAATCAGCACAATTCGTTCAATTAATAAAAAAGCGGTAAACGAGCA 1895  
Qy 84 AspThrTyrLeuGln 88  
Db 1896 GACACGTATCTGGAA 1910

## RESULT 3

US-11-179-411-21  
Sequence 21, Application US/11179411  
Publication No. US20050266491A1  
GENERAL INFORMATION:  
APPLICANT: Bryan, Bruce  
APPLICANT: Szent-Gyorgyi, Christopher  
APPLICANT: Szczepaniak, William  
TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING THEM  
TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH THROUGHPUT SCREENING AND NOVELTY ITEMS  
FILE REFERENCE: 24729-128  
CURRENT APPLICATION NUMBER: US/11/179,411  
CURRENT FILING DATE: 2005-07-12  
PRIOR APPLICATION NUMBER: US/09/808,898  
PRIOR FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: 60/189,691  
PRIOR FILING DATE: 2000-03-15  
PRIOR APPLICATION NUMBER: 09/277,716  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 08/757,046  
PRIOR FILING DATE: 1996-11-25  
PRIOR APPLICATION NUMBER: 08/597,274  
PRIOR FILING DATE: 1996-02-06  
PRIOR APPLICATION NUMBER: 08/908,909  
PRIOR FILING DATE: 1997-08-08  
PRIOR APPLICATION NUMBER: 08/990,103  
PRIOR FILING DATE: 1997-12-12  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: RastSEQ for Windows Version 4.0  
SEQ ID NO 21  
LENGTH: 1146  
TYPE: DNA  
ORGANISM: Gaussia  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)....(1146)  
OTHER INFORMATION: Nucleotide sequence encoding a CBD-Gaussia  
OTHER INFORMATION: luciferase fusion protein  
US-11-179-411-21

Alignment Scores:  
Pred. No.: 0.00017 Length: 1146  
Score: 113.00 Matches: 29  
Percent Similarity: 53.2% Conservative: 21  
Best Local Similarity: 30.9% Mismatches: 32  
Query Match: 24.1% Indels: 12  
DB: 12 Gaps: 4

US-09-917-376-4 (1-89) x US-11-179-411-21 (1-1146)

Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24  
Db 1 ATGTCAGTTGAATTTTACAACTCTAAACAAATCAGCAACAACTCAATTAACCAATA 60  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
Db 61 ATCAAAATTAATTAACATCTGACAGTATTAAATTTAAATGAGTAAAGTTAGATAT 120



Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
Db 121 TATTACACAAAGTGTGATACAAAGACAAACTTTCTCG---TGTGACCATGCTGTGCA 177  
Qy 65 GlyCysGlyAsn-----IleArgAlaSerPhe-----Gly 74  
Db 178 TTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAA 237  
Qy 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 238 ACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 276

## RESULT 4

US-11-175-766-21  
; Sequence 21, Application US/11175766  
; Publication No. US2005027211A1  
; GENERAL INFORMATION:  
; APPLICANT: Bryan, Bruce  
; APPLICANT: Szent-Gyorgyi, Christopher  
; APPLICANT: Szczepaniak, William  
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T  
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH  
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS  
; FILE REFERENCE: 24729-128  
; CURRENT APPLICATION NUMBER: US/11/175,766  
; PRIOR FILING DATE: 2005-07-06  
; PRIOR APPLICATION NUMBER: US/09/808,898  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: 60/189,691  
; PRIOR FILING DATE: 2000-03-15  
; PRIOR APPLICATION NUMBER: 09/277,716  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 08/757,046  
; PRIOR FILING DATE: 1996-11-25  
; PRIOR APPLICATION NUMBER: 08/597,274  
; PRIOR FILING DATE: 1996-02-06  
; PRIOR APPLICATION NUMBER: 08/908,909  
; PRIOR FILING DATE: 1997-08-08  
; PRIOR APPLICATION NUMBER: 08/990,103  
; PRIOR FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 21  
; LENGTH: 1146  
; TYPE: DNA  
; ORGANISM: Gausia  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1146)  
; OTHER INFORMATION: Nucleotide sequence encoding a CBD-Gausia  
; OTHER INFORMATION: luciferase fusion protein  
US-11-175-766-21

## Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00017 | Length:       | 1146 |
| Score:                 | 113.00  | Matches:      | 29   |
| Percent Similarity:    | 53.2%   | Conservative: | 21   |
| Best Local Similarity: | 30.9%   | Mismatches:   | 32   |
| Query Match:           | 24.1%   | Indels:       | 12   |
| DB:                    | 12      | Gaps:         | 4    |

US-09-917-376-4 (1-89) x US-11-175-766-21 (1-1146)

Qy 5 VallysValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
Db 1 ATGTCAGTTGAATTTTAACTCTAACAATCAGCAACAACTCAATTACACCAATA 60  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
Db 61 ATCAAAATTTACTACATCTCGACAGTGATTTAAATTTAAATGACGTAAGTTAGATAT 120  
Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64

Db 121 TATTACACAAAGTGTGATACAAAGACAAACTTTCTCG---TGTGACCATGCTGTGCA 177  
Qy 65 GlyCysGlyAsn-----IleArgAlaSerPhe-----Gly 74  
Db 178 TTATTAGGAATAGCTATGTTGATAACACTAGCAAGTGACAGCAAACTTCGTTAAAGAA 237  
Qy 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 238 ACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 276

## RESULT 5

US-10-517-939-343  
; Sequence 343, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Esteghlalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 343  
; LENGTH: 2217  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
US-10-517-939-343

## Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00159 | Length:       | 2217 |
| Score:                 | 108.00  | Matches:      | 27   |
| Percent Similarity:    | 53.6%   | Conservative: | 18   |
| Best Local Similarity: | 32.1%   | Mismatches:   | 31   |
| Query Match:           | 23.0%   | Indels:       | 8    |
| DB:                    | 8       | Gaps:         | 4    |

US-09-917-376-4 (1-89) x US-10-517-939-343 (1-2217)

Qy 7 ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsn-----GlnIleLysProGly 24  
Db 1771 CTTCAAGTACCGC-----TCCGCTGATGAGATAACAACATATCAATGAAGCCTCAG 1821  
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
Db 1822 TTCACGATCAAGAAGCAGGCAAGTGCCTCCCGTTAAGCGAGTGACGATCCGCTAC 1881  
Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet 64  
Db 1882 TATTTCACCGCGAG---AGCACGACGCGGTGGATACAGGATCGACTGGGCCCAATTC 1938  
Qy 65 GlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAsp 84  
Db 1939 GGAGCAGACGATGTCACGACGACGCTC-----GTTCCGCCATCCGATGCCGGCGGCAC 1992  
Qy 85 ThrTyrLeuGln 88  
Db 1993 GCCTATGTCGAG 2004

## RESULT 6

```
US-11-052-554A-660
; Sequence 660, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 660
; LENGTH: 15876
; TYPE: DNA
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-660

Alignment Scores:
Pred. No.: 1.15e+03 Length: 15876
Score: 70.00 Matches: 29
Percent Similarity: 50.6% Conservative: 15
Best Local Similarity: 33.3% Mismatches: 19
Query Match: 14.9% Indels: 24
DB: 12 Gaps: 7

US-09-917-376-4 (1-89) x US-11-052-554A-660 (1-15876)
Qy 17 GlyAsp-----AsnGlnIleLysProGlyLeuGlnLeuValAsnThrGlySerSer 33
Db 2542 GGTGATGATGTGATCAACGCCGTCGAAAGGCGAAGCGTGTGCTGTAAGCGCGACGCC 2601

Qy 34 Ser---ValAspLeu---SerThrValThrValArgTyrTrpPheThrArgAspGlyGly 51
Db 2602 AGCGGTGTGCGACGCGGCGGACGCGTGCACCTTT-----GGCGGC 2646

Qy 52 -----SerSerThrLeuValTyrAsnCysAspTrp----- 61
Db 2647 AAAAATATACACACACAGTGAAGTAAACGCTAGCTGGACGGTGAATGTGCGCGCTGCC 2706

Qy 62 -----AlaAlaMet-----GlyCysGlyAsnIleArgAlaSerPheGlySerValAsn 77
Db 2707 GATCTCGTGGCTACCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2766

Qy 78 ProAlaThrProThrAlaAsp 84
Db 2767 GGCAACAGCGCCCGCCAGCCGAT 2787

RESULT 7
US-09-925-065A-353393
; Sequence 353393, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
```

```
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 353393
; LENGTH: 617
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-353393

Alignment Scores:
Pred. No.: 31.8 Length: 617
Score: 69.50 Matches: 19
Percent Similarity: 44.8% Conservative: 11
Best Local Similarity: 28.4% Mismatches: 20
Query Match: 14.8% Indels: 17
DB: 6 Gaps: 2

US-09-917-376-4 (1-89) x US-09-925-065A-353393 (1-617)
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 214 TTGAGAGCAGTTAAGGGAGGGTCAAGATCTTGTAGCCTCAGCTCCATGACTCTAGGCCA 273

Qy 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla--- 63
Db 274 TGGTTT-----CTAATCTTATGCTAGTTTCTTGGTCTGCTCCC 312

Qy 64 -----MetGlyCysGlyAsnIleArgAlaSerPheGly 74
Db 313 CAGCAAGAGTAAAGCCTATCTTTAAGAGGGCGCTGTACCATCTTTGTTTATAGACTATAA 372

Qy 75 SerValAsnProAlaThrPro 81
Db 373 ACTATAAACCCAGCTCTCTCCA 393

RESULT 8
US-10-750-185-31182
; Sequence 31182, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31182
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Bovine 19866880926529
US-10-750-185-31182

Alignment Scores:
Pred. No.: 81.8 Length: 1084
Score: 68.50 Matches: 25
Percent Similarity: 40.2% Conservative: 12
Best Local Similarity: 27.2% Mismatches: 32
Query Match: 14.6% Indels: 23
DB: 8 Gaps: 5

US-09-917-376-4 (1-89) x US-10-750-185-31182 (1-1084)
Qy 16 ProGlyAspAsnGlnIleLysProGlyLeuGlnLeuVal-----AsnThr 30
Db 738 CCAGGCACCAAC-----CCCAATTCTCACCTGGTGTCTCTCTCTCTCTCTCTCTCTCTG 788
```



; PRIOR FILING DATE: 2005-05-26  
 ; NUMBER OF SEQ ID NOS: 362830  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 487  
 ; LENGTH: 4952  
 ; TYPE: DNA  
 ; ORGANISM: Rattus norvegicus  
 US-11-136-527-487

Alignment Scores:  
 Pred. No.: 470 Length: 4952  
 Score: 68.50 Matches: 22  
 Percent Similarity: 44.6% Conservative: 7  
 Best Local Similarity: 33.8% Mismatches: 25  
 Query Match: 14.6% Indels: 11  
 DB: 12 Gaps: 3

US-09-917-376-4 (1-89) x US-11-136-527-487 (1-4952)

Qy 31 GlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGly 50  
 Db 1413 GGGACCCAGAACAGACCTAGAGTCTAGTGCAGCATTATACACTCAAGGCAGATGGA 1472  
 Qy 51 -----GlySerSer-----ThrLeuValTyrAsnCysAsp----- 60  
 Db 1473 ACCAGTGTCCAGTGTCTCTTCGGAACAACAGTCATCTACACGTGTGAGTTGTCGACG 1532  
 Qy 61 ---TipAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 79  
 Db 1533 GTCTACGGACCAAGCAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGT 1592  
 Qy 80 ThrProThrAlaAsp 84  
 Db 1593 ACAATAACCCCGGAC 1607

## RESULT 12

US-10-517-939-329  
 ; Sequence 329, Application US/10517939  
 ; Publication No. US2006003433A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steer, Brian  
 ; APPLICANT: Callen, Walter  
 ; APPLICANT: Healey, Shaun  
 ; APPLICANT: Hazlewood, Geoff  
 ; APPLICANT: Wu, Di  
 ; APPLICANT: Blum, David  
 ; APPLICANT: Esteghalian, Alireza  
 ; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
 ; FILE REFERENCE: 564462007901  
 ; CURRENT APPLICATION NUMBER: US/10/517,939  
 ; PRIOR FILING DATE: 2004-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/US03/19153  
 ; PRIOR FILING DATE: 2003-06-16  
 ; PRIOR APPLICATION NUMBER: 60/389,299  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 380  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 329  
 ; LENGTH: 2268  
 ; TYPE: DNA  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: Obtained from an environmental sample.  
 US-10-517-939-329

Alignment Scores:  
 Pred. No.: 222 Length: 2268  
 Score: 68.00 Matches: 31  
 Percent Similarity: 37.9% Conservative: 8  
 Best Local Similarity: 30.1% Mismatches: 40  
 Query Match: 14.5% Indels: 24  
 DB: 8 Gaps: 5

US-09-917-376-4 (1-89) x US-10-517-939-329 (1-2268)  
 Qy 2 SerGlyGlyValLysValGlnTyrLys---AsnAsnAspSerAlaProGlyAspAsnGln 20  
 Db 1951 TCAGGGCGCTGCACCGTCGCTACGCCATCACCACGACTGGGGCAGCGGT----- 2001  
 Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
 Db 2002 TTCACCGGCAACGTTACCTCACCATACTGGCGGAAGCGCC-----CTCAACGGCTGG 2055  
 Qy 41 ThrValArgTyrTrpPheThrArgAsp----- 49  
 Db 2056 ACCCTGGCTATGCTTTCCGCGCAATCAAAACCATCAGCAACGCTGGAAACGGAACGGCC 2115  
 Qy 50 -----GlyGlySerSerThrLeuValTyrAsnCysAspTyrAlaAlaMetGlyCysGly 67  
 Db 2116 GTTCAGTCCGCGCAGCAGCTCAGCGTCACCAACGCCGTTGGAATGGCAGCTGCGGCC 2175  
 Qy 68 AsnIleArgAlaSerPheGly-----SerValAsnProAlaThrProThr 82  
 Db 2176 AACGTCTCCGCCAGCTTTGGCTTCCAGCGAGCTACAGCGGCAATAACAGCGTCCCTGCC 2235  
 Qy 83 AlaAspThr 85  
 Db 2236 AGCTTTACG 2244

## RESULT 13

US-10-995-561-13307/c  
 ; Sequence 13307, Application US/10995561  
 ; Publication No. US20050272054A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele et al.  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
 ; FILE REFERENCE: CL001559  
 ; CURRENT APPLICATION NUMBER: US/10/995,561  
 ; CURRENT FILING DATE: 2004-11-24  
 ; NUMBER OF SEQ ID NOS: 85702  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13307  
 ; LENGTH: 35344  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-995-561-13307

Alignment Scores:  
 Pred. No.: 5,23e+03 Length: 35344  
 Score: 68.00 Matches: 16  
 Percent Similarity: 57.1% Conservative: 4  
 Best Local Similarity: 45.7% Mismatches: 11  
 Query Match: 14.5% Indels: 4  
 DB: 8 Gaps: 1

US-09-917-376-4 (1-89) x US-10-995-561-13307 (1-35344)

Qy 51 GlySerSerThrLeuValTyr-----AsnCysAspTyrAlaAlaMetGlyCys 66  
 Db 23968 GGTAAGACCAACACTTCCCTATTCACCTTGAGGTCACTGCTGGCGGCTTTGTTTGGC 23909  
 Qy 67 GlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81  
 Db 23908 TTTAGAGCTAGGGCATCATTTACCACCTTCTTCCCTCTAGTCTCT 23864

## RESULT 14

US-09-925-065A-228751/c  
 ; Sequence 228751, Application US/09925065A  
 ; Publication No. US20040181048A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

```
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228751
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-228751

Alignment Scores:
Pred. No.: 51      Length: 556
Score: 67.50      Matches: 25
Percent Similarity: 37.6%      Conservative: 7
Best Local Similarity: 29.4%      Mismatches: 20
Query Match: 14.4%      Indels: 33
DB: 6      Gaps: 4

US-09-917-376-4 (1-89) x US-09-925-065A-228751 (1-556)
QY 23 ProGlyLeuGlnLeuValAsnThrGlySer----- 32
Db 238 CCGTGGACTCCACCTTGTGCATACAGGCAGTTAAGGAATGCTTTCGTCGACAAAGACTTTT 179
QY 33 -----SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArg 48
Db 178 ATTTAGCTTGTAGTTCTCTCGAATTTGTTTGGAGATTGAGTTACTGCTGTAATAATATGG 119
QY 49 AspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet----- 64
Db 118 GACCCGGGG-----TAGGACTGGTGGGGCTCTGCATCCCTTCC 80
QY 65 -----GlyCysGlyAsnIleArg---AlaSerPheGlySerVal 76
Db 79 TCCGAGGGGCCCTGGGAGAGGGGTGGGACTGCCCGCGTGGTGTTCCTCCCCCAATT 20
QY 77 AsnProAlaThrPro 81
Db 19 TCCTGTTGGACCCCT 5

RESULT 15
US-11-072-512-279/c
; Sequence 279, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 279
; LENGTH: 2564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-279

Alignment Scores:
Pred. No.: 296      Length: 2564
Score: 67.50      Matches: 23
Percent Similarity: 42.1%      Conservative: 9
Best Local Similarity: 30.3%      Mismatches: 27
Query Match: 14.4%      Indels: 17
DB: 9      Gaps: 3

US-09-917-376-4 (1-89) x US-11-072-512-279 (1-2564)
QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 454 CCTCCGAAGCGATGGCTGCACACTCGCTCCAGCAGCTTGTCCCGCAGCTCTCGAATACC 395
QY 43 ArgTyrTrpPheThr----- 47
Db 394 GACTCGTGGAACTCCAGCGCGCGTCCGTCGAGGTCCCAAGAGGGGAGGGCTGACTGA 335
QY 48 ArgAspGlyGlySerSerThrLeu---ValTyrAsnCysAspTrpAlaAlaMetGlyCys 66
Db 334 AGAGATGGCAGCAGCACACCAATTCTCTGTCTGGAAGTCTCGATGCTTGTAGCGGCTCC 275
QY 67 GlyAsn---IleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 274 GTGCAGTTGGTCAGGGTCTCTTCAGGTCCTCGCCCGTTGGCCACGCC 227

Search completed: March 2, 2006, 13:23:22
Job time : 103.506 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 13:55:49 ; Search time 5.0715 Seconds  
(without alignments)  
1434.576 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVQVQKNDSPGNDQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5 COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6 COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description       |
|------------|-------|-------------|--------|----|-------------------|
| 1          | 195.5 | 41.7        | 616    | 2  | US-09-136-574A-47 |
| 2          | 194.5 | 41.5        | 1751   | 2  | US-09-136-574A-44 |
| 3          | 191.5 | 40.8        | 1426   | 2  | US-09-136-574A-43 |
| 4          | 172   | 36.7        | 700    | 1  | US-07-862-588B-2  |
| 5          | 165.5 | 35.3        | 551    | 1  | US-09-033-537A-1  |
| 6          | 165   | 35.2        | 167    | 4  | PCT-US95-13813-9  |
| 7          | 165   | 35.2        | 476    | 2  | US-09-339-159B-4  |
| 8          | 165   | 35.2        | 493    | 2  | US-09-198-955A-10 |
| 9          | 165   | 35.2        | 493    | 2  | US-09-198-955A-12 |
| 10         | 165   | 35.2        | 493    | 2  | US-09-694-531-12  |
| 11         | 165   | 35.2        | 493    | 2  | US-09-670-141-10  |
| 12         | 165   | 35.2        | 493    | 2  | US-10-072-152-12  |
| 13         | 155.5 | 33.2        | 1352   | 2  | US-09-784-554B-2  |
| 14         | 150.5 | 32.1        | 1350   | 2  | US-09-784-554B-4  |
| 15         | 120.5 | 25.7        | 531    | 1  | US-07-862-588B-7  |
| 16         | 114   | 24.3        | 162    | 1  | US-08-048-164A-2  |
| 17         | 114   | 24.3        | 162    | 1  | US-08-460-462-2   |
| 18         | 114   | 24.3        | 162    | 1  | US-08-460-457-2   |
| 19         | 114   | 24.3        | 162    | 1  | US-08-460-458-2   |
| 20         | 114   | 24.3        | 162    | 1  | US-08-460-455-2   |
| 21         | 114   | 24.3        | 162    | 1  | US-08-330-394A-2  |
| 22         | 114   | 24.3        | 163    | 2  | US-09-006-636-7   |
| 23         | 114   | 24.3        | 163    | 2  | US-09-006-632-7   |
| 24         | 114   | 24.3        | 163    | 2  | US-09-325-274-7   |
| 25         | 113   | 24.1        | 382    | 2  | US-09-277-716-22  |
| 26         | 113   | 24.1        | 382    | 2  | US-09-609-161B-22 |
| 27         | 112   | 23.9        | 154    | 1  | US-08-330-394A-29 |

28 112 23.9 156 1 US-08-330-394A-22 Sequence 22, Appl  
29 72.5 15.5 272 2 US-09-902-540-11056 Sequence 11056, A  
30 68.5 14.6 500 2 US-10-104-047-2825 Sequence 2825, Ap  
31 65.5 14.0 1566 2 US-09-581-472B-2 Sequence 2, Appli  
32 65 13.9 474 2 US-09-248-796A-25524 Sequence 25524, A  
33 64 13.6 143 2 US-09-301-593-26 Sequence 26, Appl  
34 64 13.6 428 2 US-09-118-319-5 Sequence 5, Appli  
35 64 13.6 464 1 US-08-353-400-36 Sequence 36, Appl  
36 64 13.6 472 2 US-09-301-593-30 Sequence 30, Appl  
37 63.5 13.5 1785 2 US-09-341-587-3 Sequence 3, Appli  
38 63 13.4 453 2 US-09-301-593-18 Sequence 18, Appl  
39 63 13.4 472 2 US-09-301-593-43 Sequence 43, Appl  
40 63 13.4 718 2 US-09-328-352-4640 Sequence 4640, Ap  
41 63 13.4 1133 2 US-09-902-540-12243 Sequence 12243, A  
42 63 13.4 1581 2 US-09-110-517-2 Sequence 2, Appli  
43 62.5 13.3 288 2 US-09-423-439-38 Sequence 38, Appl  
44 62.5 13.3 389 2 US-09-902-540-14347 Sequence 14347, A  
45 62.5 13.3 445 1 US-08-353-400-33 Sequence 33, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-136-574A-47  
; Sequence 47, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 616 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-09-136-574A-47

Query Match 41.7%; Score 195.5; DB 2; Length 616;  
Best Local Similarity 43.2%; Pred. No. 3.4e-13;  
Matches 38; Conservative 15; Mismatches 34; Indels 1; Gaps 1;

QY 1 VSGGVKVKYKNDSPAGNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 60  
DB 1 MCGVKVLYKKNETSASTGSRPWFKIVNGGSSVDLSRVKIRYWTVDGKPKQSAV-CD 59  
QY 61 WAAMCGNIRASFGSVNPTADTYLQ 88  
DB 60 WAQIGASNVTFNFKVLSGSGVSGADYYLE 87

RESULT 2

US-09-136-574A-44  
; Sequence 44, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 1751 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 44:

US-09-136-574A-44

Query Match 41.5%; Score 194.5; DB 2; Length 1751;  
Best Local Similarity 44.7%; Pred. No. 1.6e-12;  
Matches 38; Conservative 14; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63  
DB 413 GLKVLKKNETSASTGSRPWFKIVNGGSSVDLSRVKIRYWTVDGKPKQSAV-CD 471

RESULT 4

DB 678 GVKVLYKKNETSASTGSRPWFKIVNGGSSVDLSRVKIRYWTVDGKPKQSAV-CD 736

QY 64 MCGNIRASFGSVNPTADTYLQ 88  
DB 737 IGASNVTFNFKVLSGSGVSGADYYLE 761

RESULT 3

US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.  
Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1426 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e  
SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 40.8%; Score 191.5; DB 2; Length 1426;  
Best Local Similarity 43.5%; Pred. No. 2.6e-12;  
Matches 37; Conservative 15; Mismatches 32; Indels 1; Gaps 1;

QY 4 GVKVQYKNDSPAGNQIKPGLQLVNTGSSVDLSRVKIRYWTVDGKPKQSAV-CD 63  
DB 413 GLKVLKKNETSASTGSRPWFKIVNGGSSVDLSRVKIRYWTVDGKPKQSAV-CD 471

QY 64 MCGNIRASFGSVNPTADTYLQ 88  
DB 472 IGASNVTFNFKVLSGSGVSGADYYLE 496

```
US-07-862-588B-2
; Sequence 2, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-2

Query Match 36.7%; Score 172; DB 1; Length 700;
Best Local Similarity 44.9%; Pred. No. 1.5e-10;
Matches 40; Conservative 18; Mismatches 27; Indels 4; Gaps 3;

QY 1 VSGGVQVQKND-SAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTDGGSSTLVYNC 59
Db 549 VNSDLVQVQKGRNATDNQIKPHFNQKTSVDLSLTRYFTKD-SSAAMNGWI 607
QY 60 DWAAMCGNIRASFGSNVNPATPTADTYLQ 88
Db 608 DWAUKGSNIQSFQNHGA--DSDTYAE 634

RESULT 5
US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Merete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sch lein, Martin
; TITLE OF INVENTION: Prevention Of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
```

```
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/033,537A
FILING DATE: 02-MAR-1998
CLASSIFICATION: 008
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 0993/95
FILING DATE: 08-SEP-1995
APPLICATION NUMBER: PCT/DK96/00364
FILING DATE: 03-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Green, Reza
REGISTRATION NUMBER: 38,475
REFERENCE/DOCKET NUMBER: 4492.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 551 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-033-537A-1

Query Match 35.3%; Score 165.5; DB 1; Length 551;
Best Local Similarity 37.9%; Pred. No. 6e-10;
Matches 33; Conservative 22; Mismatches 29; Indels 3; Gaps 2;

QY 2 SGGVKVQYKND-SAPGDNQIKPGLQVNTGSSVDLSTVTVRYWFTDGGSSTLVYNC 61
Db 402 TGNLVQVQKVDTSATDNQMKPSFNKNKNTTPVNLGSLKLYFTKD-GTADMSASPDW 460
QY 62 AAMCGNIRASFGSNVNPATPTADTYLQ 88
Db 461 AQIGASNVSAAP--ANFTGSNTDTIYE 485

RESULT 6
PCT-US95-13813-9
; Sequence 9, Application PC/TUS9513813
; GENERAL INFORMATION:
; APPLICANT: Yeda Research and Development Co. Ltd.
; APPLICANT: Ramot University Authority for Applied
; APPLICANT: Research and Industrial Development Ltd.
; APPLICANT: Technion Research and Development Foundation Ltd.
; APPLICANT: Bayer, Edward A.
; APPLICANT: Morag, Ely
; APPLICANT: Wilchek, Meir
; APPLICANT: Lamed, Raphael
; APPLICANT: Shoham, Yuval
; TITLE OF INVENTION: MODIFIED CELLULOSE-BINDING DOMAIN (CBD)
; TITLE OF INVENTION: PROTEINS AND USE THEREOF
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brody and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13813  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: BAYER=3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 167 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-13813-9

Query Match 35.2%; Score 165; DB 4; Length 167;  
Best Local Similarity 36.1%; Pred. No. 1.6e-10;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVYWFTRDGGSSTLVYVNC 60  
DB 5 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYTVYVDGQDQTFW-CD 63  
QY 61 WAAM-----GCGNIRASFGSVNPTPTADTVLQ 88  
DB 64 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 100

## RESULT 7

US-09-339-159B-4  
Sequence 4, Application US/09339159B  
Patent No. 6566114  
GENERAL INFORMATION:  
APPLICANT: Kauppinen, Markus  
APPLICANT: Schulein, Martin  
APPLICANT: Schnorr, Kirk  
APPLICANT: Andersen, Lene  
APPLICANT: Bjornvad, Mads  
TITLE OF INVENTION: No. 6566114el Mannanases  
FILE REFERENCE: 5440.204-US  
CURRENT APPLICATION NUMBER: US/09/339,159B  
CURRENT FILING DATE: 1999-06-24  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Bacillus  
US-09-339-159B-4

Query Match 35.2%; Score 165; DB 2; Length 476;  
Best Local Similarity 36.1%; Pred. No. 5.7e-10;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVYWFTRDGGSSTLVYVNC 60  
DB 314 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYTVYVDGQDQTFW-CD 372  
QY 61 WAAM-----GCGNIRASFGSVNPTPTADTVLQ 88  
DB 373 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 409

## RESULT 8

US-09-198-956-10  
Sequence 10, Application US/09198956  
Patent No. 6165769  
GENERAL INFORMATION:  
APPLICANT: Andersen, Lene N.  
APPLICANT: Schulein, Martin  
APPLICANT: Lange, Niels Erik K.  
APPLICANT: Bjornvad, Mads E.  
APPLICANT: Schnorr, Kirk  
TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus  
FILE REFERENCE: 5377.200-US  
CURRENT APPLICATION NUMBER: US/09/198,956  
CURRENT FILING DATE: 1998-11-24  
EARLIER APPLICATION NUMBER: 1344/97  
EARLIER FILING DATE: 1997-11-24  
EARLIER APPLICATION NUMBER: 60/067,240  
EARLIER FILING DATE: 1997-12-02  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10  
LENGTH: 493  
TYPE: PRT  
ORGANISM: Bacillus licheniformis  
US-09-198-956-10

Query Match 35.2%; Score 165; DB 2; Length 493;  
Best Local Similarity 36.1%; Pred. No. 5.9e-10;  
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;  
QY 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVYWFTRDGGSSTLVYVNC 60  
DB 331 VSGNLKVEFYNSPDDTNSINPQFKVTNTGSSAIDLKLTLYRYTVYVDGQDQTFW-CD 389  
QY 61 WAAM-----GCGNIRASFGSVNPTPTADTVLQ 88  
DB 390 HAAIGSGSYNGITSNVKGTFFVKMSSSTNNADTYLE 426

## RESULT 9

US-09-198-955A-12  
Sequence 12, Application US/09198955A  
Patent No. 6187580  
GENERAL INFORMATION:  
APPLICANT: Andersen, Lene N.  
APPLICANT: Schulein, Martin  
APPLICANT: Lange, Niels E.  
APPLICANT: Bjornvad, Mads E.  
APPLICANT: Moller, Soren  
APPLICANT: Glad, Sanne O. S.  
APPLICANT: Kauppinen, Markus S.  
APPLICANT: Schnorr, Kirk  
APPLICANT: Kongsbak, Lars  
TITLE OF INVENTION: No. 6187580el Pectate Lyases  
FILE REFERENCE: 5378.200-US  
CURRENT APPLICATION NUMBER: US/09/198,955A  
CURRENT FILING DATE: 1998-11-24  
PRIOR APPLICATION NUMBER: 1343/97  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 1344/97  
PRIOR FILING DATE: 1997-11-24  
PRIOR APPLICATION NUMBER: 60/067,249  
PRIOR FILING DATE: 1997-12-02  
PRIOR APPLICATION NUMBER: 60/067,240  
PRIOR FILING DATE: 1997-12-02  
PRIOR APPLICATION NUMBER: 09/073,684  
PRIOR FILING DATE: 1998-05-06  
PRIOR APPLICATION NUMBER: 09/184,217  
PRIOR FILING DATE: 1998-11-02  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 493

```
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-198-955A-12

Query Match      35.2%; Score 165; DB 2; Length 493;
Best Local Similarity 36.1%; Pred. No. 5.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 331 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYYYTVDGQKQTFW-CD 389

Qy 61 WAAM-----CGCNIRASFGSVNPATPTADTYLQ 88
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 390 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 426

RESULT 10
US-09-694-531-12
; Sequence 12, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; CURRENT FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-09-694-531-12

Query Match      35.2%; Score 165; DB 2; Length 493;
Best Local Similarity 36.1%; Pred. No. 5.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 331 VSGNLKVEFYNSPDDTNSINPQKVTNTGSSAIDLSKLTLYYYTVDGQKQTFW-CD 389

Qy 61 WAAM-----CGCNIRASFGSVNPATPTADTYLQ 88
   |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 390 HAAIIGNSYNGITSNVKGTFFVKSSSTNNADTYLE 426

RESULT 11
US-09-670-141-10
; Sequence 10, Application US/09670141
```

```
; SEQ ID NO 12
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-072-152-12

Query Match      35.2%; Score 165; DB 2; Length 493;
Best Local Similarity 36.1%; Pred. No. 5.9e-10;
Matches 35; Conservative 23; Mismatches 29; Indels 10; Gaps 2;

Qy 1 VSGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCD 60
Db 331 VSGNLKVEFYNSPDTNSINPQKVTNTSSAIDSLKIRIYYTVDGQKDTFW-CD 389

Qy 61 WAAM-----CGGNIRASFGSVNPTADTYLQ 88
Db 390 HAAIGSNYSNGITSNVKGTGFKMSSSTNNADTYLE 426

RESULT 13
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017-200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match      33.2%; Score 155.5; DB 2; Length 1352;
Best Local Similarity 31.8%; Pred. No. 2.3e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1200 GLLLQYRTADTKVNDNLNLPQFQVKNKTSIPINELKIRIYYTIDGREQ-TFNCDYAT 1258

Qy 64 MCGGNIRASFGSVNPTADTYLQ 88
Db 1259 LSCSKLNGKLVKDKAATGADYYLE 1283

RESULT 14
US-09-784-554B-4
; Sequence 4, Application US/09784554B
; Patent No. 6815192
; GENERAL INFORMATION:
; APPLICANT: Schnorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schulein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLUCANASES
; FILE REFERENCE: 10017-200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-4

Query Match      32.1%; Score 150.5; DB 2; Length 1350;

Best Local Similarity 31.8%; Pred. No. 8.1e-08;
Matches 27; Conservative 21; Mismatches 36; Indels 1; Gaps 1;

Qy 4 GVKVQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDWAA 63
Db 1199 GLVLQYRTADTKVNDNLNLPQFQVKNKTSIPINELKIRIYYTIDGREQ-TFNCDYAV 1257

Qy 64 MCGGNIRASFGSVNPTADTYLQ 88
Db 1258 LSCSKLNGKLVKDKAATGADYYLE 1282

RESULT 15
US-07-862-588B-7
; Sequence 7, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linaa
; APPLICANT: Schlein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5916796o No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425-204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 531 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-862-588B-7

Query Match      25.7%; Score 120.5; DB 1; Length 531;
Best Local Similarity 32.9%; Pred. No. 5.2e-05;
Matches 28; Conservative 18; Mismatches 36; Indels 3; Gaps 2;

Qy 2 SGGVKQYKNDAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVNCDW 61
Db 432 TGNLVQYKVGDTSATDNQMKPSFNKKNKTPVNLGLKXXXXKD-GPADMSCSIDW 490

Qy 62 AAMCGGNIRASFGSVNPTADTY 86
Db 491 AQIGRTNVLLAF--ANFTGSNTDTY 513

Search completed: March 2, 2006, 13:58:00
```

Job time : 5.0715 secs

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:20:38 ; Search time 16.952 Seconds  
(without alignments)  
2169.009 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQVKNDSAPGDNQ.....IRASFGSVNPTPTADTYLQ 88

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pap:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pap:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Length | DB ID | Description         |
|------------|-------|---------|--------|-------|---------------------|
| 1          | 469   | 100.0   | 88     | 3     | US-09-917-376-5     |
| 2          | 469   | 100.0   | 88     | 4     | US-10-155-400-5     |
| 3          | 469   | 100.0   | 89     | 3     | US-09-917-376-4     |
| 4          | 469   | 100.0   | 89     | 4     | US-10-155-400-4     |
| 5          | 469   | 100.0   | 154    | 3     | US-09-917-378-4     |
| 6          | 469   | 100.0   | 762    | 3     | US-09-917-378-1     |
| 7          | 466   | 99.4    | 150    | 3     | US-09-917-384-5     |
| 8          | 466   | 99.4    | 150    | 3     | US-09-917-383-5     |
| 9          | 466   | 99.4    | 1043   | 3     | US-09-917-384-6     |
| 10         | 466   | 99.4    | 1043   | 3     | US-09-917-383-6     |
| 11         | 466   | 99.4    | 1228   | 3     | US-09-917-384-1     |
| 12         | 466   | 99.4    | 1228   | 3     | US-09-917-383-1     |
| 13         | 462   | 98.5    | 957    | 3     | US-09-917-376-1     |
| 14         | 462   | 98.5    | 957    | 4     | US-10-155-400-1     |
| 15         | 257   | 54.8    | 741    | 4     | US-10-156-761-8100  |
| 16         | 190   | 40.5    | 508    | 4     | US-10-369-493-23151 |
| 17         | 188   | 40.1    | 477    | 4     | US-10-466-208-12    |
| 18         | 188   | 40.1    | 496    | 4     | US-10-466-208-8     |
| 19         | 188   | 40.1    | 677    | 4     | US-10-433-577-35    |
| 20         | 180   | 38.4    | 1621   | 4     | US-10-185-990-10    |
| 21         | 165   | 35.2    | 167    | 5     | US-10-933-404-4     |
| 22         | 165   | 35.2    | 476    | 4     | US-10-372-054-4     |
| 23         | 165   | 35.2    | 493    | 4     | US-10-072-152-12    |
| 24         | 165   | 35.2    | 493    | 4     | US-10-655-433-12    |
| 25         | 165   | 35.2    | 599    | 3     | US-09-955-555A-29   |
| 26         | 155.5 | 33.2    | 1352   | 3     | US-09-784-5548-2    |
| 27         | 155.5 | 33.2    | 1352   | 5     | US-10-896-555-2     |

|    |       |      |      |   |                      |                    |
|----|-------|------|------|---|----------------------|--------------------|
| 28 | 150.5 | 32.1 | 1350 | 3 | US-09-784-554B-4     | Sequence 4, Appli  |
| 29 | 150.5 | 32.1 | 1350 | 5 | US-10-896-555-4      | Sequence 4, Appli  |
| 30 | 130.5 | 27.8 | 1483 | 4 | US-10-282-122A-51483 | Sequence 51483, A  |
| 31 | 114   | 24.3 | 256  | 4 | US-10-261-446-6      | Sequence 6, Appli  |
| 32 | 114   | 24.3 | 256  | 4 | US-10-261-445B-6     | Sequence 6, Appli  |
| 33 | 114   | 24.3 | 256  | 5 | US-10-782-234-6      | Sequence 6, Appli  |
| 34 | 113   | 24.1 | 382  | 3 | US-09-808-898-22     | Sequence 22, Appli |
| 35 | 105.5 | 22.5 | 183  | 4 | US-10-460-524-2      | Sequence 2, Appli  |
| 36 | 78.5  | 16.7 | 1049 | 4 | US-10-282-122A-49900 | Sequence 49900, A  |
| 37 | 75.5  | 16.1 | 346  | 4 | US-10-437-963-179234 | Sequence 179234,   |
| 38 | 75.5  | 16.1 | 618  | 4 | US-10-211-462-223    | Sequence 223, App  |
| 39 | 75.5  | 16.1 | 618  | 5 | US-10-723-860-4039   | Sequence 4039, Ap  |
| 40 | 75.5  | 16.1 | 618  | 5 | US-10-756-149-5713   | Sequence 5713, Ap  |
| 41 | 75.5  | 16.1 | 986  | 3 | US-09-747-835A-53    | Sequence 53, Appl  |
| 42 | 75.5  | 16.1 | 986  | 4 | US-10-120-604-101    | Sequence 101, App  |
| 43 | 75.5  | 16.1 | 986  | 4 | US-10-225-567A-406   | Sequence 406, App  |
| 44 | 75.5  | 16.1 | 986  | 4 | US-10-398-458-3      | Sequence 3, Appli  |
| 45 | 75.5  | 16.1 | 986  | 4 | US-10-312-312-53     | Sequence 53, Appl  |

#### ALIGNMENTS

#### RESULT 1

US-09-917-376-5  
; Sequence 5, Application US/09917376  
; Publication No. US2004003834A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 88  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; OTHER INFORMATION: Carbohydrate binding domain  
US-09-917-376-5

|                       |         |  |       |            |
|-----------------------|---------|--|-------|------------|
| Query Match           | 100.0%; | Score 469;   | DB 3; | Length 88; |
| Best Local Similarity | 100.0%; | Pred. No. 3.6e-48;                                     |       |            |
| Matches               | 88;     | Conservative   | 0;    | Mismatches |
|                       |         |  | 0;    | Indels     |
|                       |         |  |       | Gaps       |
|                       |         |  |       | 0;         |
| Qy                    | 1       | VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTFRDGSSTLVNCD | 60    |            |
| Db                    | 1       | VSGGVKQVKNDSAPGDNQIKPGLQVNTGSSVDLSTVTYVWFTFRDGSSTLVNCD | 60    |            |
| Qy                    | 61      | WAAMCGNIRASFGSVNPTPTADTYLQ                             | 88    |            |
| Db                    | 61      | WAAMCGNIRASFGSVNPTPTADTYLQ                             | 88    |            |

#### RESULT 2

US-10-155-400-5  
; Sequence 5, Application US/10155400  
; Publication No. US2003010898A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400

```
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-5

Query Match      100.0%; Score 469; DB 4; Length 88;
Best Local Similarity 100.0%; Pred. No. 3.6e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
DB 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 3
US-09-917-376-4
; Sequence 4, Application US/09917376
; Publication No. US20040038344A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.4US01
; CURRENT APPLICATION NUMBER: US/09/917.376
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-09-917-376-4

Query Match      100.0%; Score 469; DB 3; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.7e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
DB 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 4
US-10-155-400-4
; Sequence 4, Application US/10155400
; Publication No. US20030108988A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: NREL 01-36A
; CURRENT APPLICATION NUMBER: US/10/155,400
; CURRENT FILING DATE: 2002-10-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (89)
; OTHER INFORMATION: Any amino acid
; OTHER INFORMATION: Carbohydrate binding domain
US-10-155-400-4

Query Match      100.0%; Score 469; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 3.7e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
DB 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 5
US-09-917-378-4
; Sequence 4, Application US/09917378
; Publication No. US20030119093A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Carbohydrate
; OTHER INFORMATION: binding domain
US-09-917-378-4

Query Match      100.0%; Score 469; DB 3; Length 154;
Best Local Similarity 100.0%; Pred. No. 7.1e-48;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60
DB 1 VSGGVKQVQYKKNDSAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGGNIRASFGSVNPATPTADTYLQ 88

RESULT 6
US-09-917-378-1
```

```

; Sequence 1, Application US/09917378
; Publication No. US2003011903A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40197.7US01
; CURRENT APPLICATION NUMBER: US/09/917,378
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Acidothermus cellulolyticus
; US-09-917-378-1

Query Match      100.0%; Score 469; DB 3; Length 762;
Best Local Similarity 100.0%; Pred. No. 4.9e-47;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 60
DB 455 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 514

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542

RESULT 7
US-09-917-384-5
; Sequence 5, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
; US-09-917-384-5

Query Match      99.4%; Score 466; DB 3; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 60
DB 455 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 514

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 515 WAAMCGNIRASFGSVNPATPTADTYLQ 542

RESULT 8
US-09-917-383-5
; Sequence 5, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
; US-09-917-383-5

Query Match      99.4%; Score 466; DB 3; Length 150;
Best Local Similarity 98.9%; Pred. No. 1.6e-47;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 60
DB 455 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 60

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88
DB 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88

RESULT 9
US-09-917-384-6
; Sequence 6, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; TITLE OF INVENTION: CELLULOLYTICUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
; OTHER INFORMATION: Guxa
; US-09-917-384-6

Query Match      99.4%; Score 466; DB 3; Length 1043;
Best Local Similarity 98.9%; Pred. No. 1.7e-46;
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 60
DB 477 VSGGVKQVYKNNDSAPGDNQIKPGLQLVNTGSSVDLSVTTVRYWFTTRDGGSSTLVYNCD 536
```

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

## RESULT 10

US-09-917-383-6  
; Sequence 6, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1043  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-6

Query Match 99.4%; Score 466; DB 3; Length 1043;  
Best Local Similarity 98.9%; Pred. No. 1.7e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGKVOYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
DB 477 VSGGLKVQYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 536  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 537 WAAMCGNIRASFGSVNPATPTADTYLQ 564

## RESULT 11

US-09-917-384-1  
; Sequence 1, Application US/09917384  
; Publication No. US20030096342A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,384  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-384-1

Query Match 99.4%; Score 466; DB 3; Length 1228;  
Best Local Similarity 98.9%; Pred. No. 2e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGKVOYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
DB 584 VSGGLKVQYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 643  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

## RESULT 12

US-09-917-383-1  
; Sequence 1, Application US/09917383  
; Publication No. US20030104522A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40170.6US01  
; CURRENT APPLICATION NUMBER: US/09/917,383  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Segment of  
; OTHER INFORMATION: Guxa  
US-09-917-383-1

Query Match 99.4%; Score 466; DB 3; Length 1228;  
Best Local Similarity 98.9%; Pred. No. 2e-46;  
Matches 87; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGKVOYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 60  
DB 584 VSGGLKVQYKXNDKAPGDNQIKPGLQLVNTGSSVDLSTVTVRYWFTRDGSSSTLVYVNC 643  
QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
DB 644 WAAMCGNIRASFGSVNPATPTADTYLQ 671

## RESULT 13

US-09-917-376-1  
; Sequence 1, Application US/09917376  
; Publication No. US20040038334A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS  
; FILE REFERENCE: 40197.4US01  
; CURRENT APPLICATION NUMBER: US/09/917,376  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid



US-09-917-376-1

Query Match 98.5%; Score 462; DB 3; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.5e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSSTLVYNC 60  
|||||  
Db 869 VSGGVKQYKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSSTLVYNC 928  
|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||

Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956  
|||||

## RESULT 14

US-10-155-400-1  
; Sequence 1, Application US/10155400  
; Publication No. US20030108988A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS  
; TITLE OF INVENTION: CELLULOGLYTICUS  
; FILE REFERENCE: NREL 01-36A  
; CURRENT APPLICATION NUMBER: US/10/155,400  
; CURRENT FILING DATE: 2002-10-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 957  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (957)  
; OTHER INFORMATION: Any amino acid  
US-10-155-400-1

Query Match 98.5%; Score 462; DB 4; Length 957;  
Best Local Similarity 97.7%; Pred. No. 4.5e-46;  
Matches 86; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VSGGVKQYKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSSTLVYNC 60  
|||||  
Db 869 VSGGVKQYKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSSTLVYNC 928  
|||||

QY 61 WAAMCGNIRASFGSVNPATPTADTYLQ 88  
|||||

Db 929 WAAIGCGNIRASFGSVNPATPTADTYLQ 956  
|||||

## RESULT 15

US-10-156-761-8100  
; Sequence 8100, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8100  
; LENGTH: 741  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8100

Query Match 54.8%; Score 257; DB 4; Length 741;  
Best Local Similarity 56.3%; Pred. No. 1.2e-21;  
Matches 49; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 2 SGGVKVQYKNDSPAGDNOIKPGLQVNTGSSVDLSTVTYRYWFTDGGSSSTLVYNC 61  
|||||  
Db 591 SGGKVLVYKNDSSATDNAIRPGLRVNTGSGSLDLSKVTRYFTTRDSSPTVNAWCDY 650  
|||||

QY 62 AAMGCGNIRASFGSVNPATPTADTYLQ 88  
|||||

Db 651 AAVGCSNVSLKVVPVLTTPVPGADAYLE 677  
|||||

Search completed: March 2, 2006, 14:26:55  
Job time: 16.952 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2006, 14:27:20 ; Search time 1.78442 Seconds  
(without alignments)  
986.322 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKVOYKNDSPGDNQ.....IRASFGSVNPATPTADTYLQ 88

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 135339 seqs, 20000136 residues

Total number of hits satisfying chosen parameters: 135339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:\*

1: /cgn2\_6/ptodata/2/pubpaa/US08 NEW PUB.pap.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06 NEW PUB.pap.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US07 NEW PUB.pap.\*  
4: /cgn2\_6/ptodata/2/pubpaa/PCT NEW PUB.pap.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US09 NEW PUB.pap.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US10 NEW PUB.pap.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US11 NEW PUB.pap.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US60 NEW PUB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description         |
|------------|-------|---------------|--------|-------|---------------------|
| 1          | 163   | 34.8          | 499    | 6     | US-10-517-939-328   |
| 2          | 157   | 33.5          | 534    | 6     | US-10-510-386-230   |
| 3          | 113   | 24.1          | 382    | 7     | US-11-179-411-22    |
| 4          | 113   | 24.1          | 382    | 7     | US-11-175-768-22    |
| 5          | 108   | 23.0          | 738    | 6     | US-10-517-939-344   |
| 6          | 72    | 15.4          | 210    | 7     | US-11-170-653-51    |
| 7          | 70    | 14.9          | 5291   | 7     | US-11-052-554A-281  |
| 8          | 68.5  | 14.6          | 500    | 7     | US-11-072-512-2825  |
| 9          | 68    | 14.5          | 755    | 6     | US-10-517-933-330   |
| 10         | 64.5  | 13.8          | 389    | 7     | US-11-207-626A-26   |
| 11         | 62    | 13.2          | 389    | 7     | US-11-207-626A-20   |
| 12         | 61    | 13.0          | 350    | 6     | US-10-517-939-14    |
| 13         | 61    | 13.0          | 1889   | 7     | US-11-102-476-46    |
| 14         | 60.5  | 12.9          | 806    | 7     | US-11-098-686-10278 |
| 15         | 60.5  | 12.8          | 983    | 7     | US-11-087-099-7483  |
| 16         | 60    | 12.8          | 86     | 7     | US-11-102-476-24    |
| 17         | 60    | 12.8          | 91     | 7     | US-11-102-476-9     |
| 18         | 60    | 12.8          | 919    | 7     | US-11-074-176-284   |
| 19         | 60    | 12.8          | 1225   | 7     | US-11-102-476-2     |
| 20         | 60    | 12.8          | 1424   | 7     | US-11-102-476-4     |
| 21         | 59.5  | 12.7          | 217    | 6     | US-10-454-437-82    |
| 22         | 59    | 12.6          | 111    | 7     | US-11-072-512-2820  |
| 23         | 59    | 12.6          | 124    | 6     | US-10-771-257-19    |
| 24         | 59    | 12.6          | 124    | 7     | US-11-127-677-19    |
| 25         | 59    | 12.6          | 182    | 7     | US-11-170-653-49    |

26 59 12.6 208 7 US-11-214-413-32 Sequence 32, Appl  
27 59 12.6 381 7 US-11-087-099-3485 Sequence 3485, Ap  
28 59 12.6 498 7 US-11-131-479-76 Sequence 76, Appl  
29 58.5 12.5 211 7 US-11-170-653-50 Sequence 50, Appl  
30 58.5 12.5 228 7 US-11-170-653-39 Sequence 39, Appl  
31 58.5 12.5 248 7 US-11-054-515-1790 Sequence 1790, Ap  
32 58.5 12.5 327 7 US-11-064-246-15 Sequence 15, Appl  
33 58.5 12.5 327 7 US-11-064-246-16 Sequence 16, Appl  
34 58.5 12.5 345 7 US-11-207-626A-37 Sequence 37, Appl  
35 58 12.4 121 7 US-11-107-028-35 Sequence 35, Appl  
36 58 12.4 130 7 US-11-188-187A-3 Sequence 3, Appl  
37 58 12.4 140 7 US-11-183-205-62 Sequence 62, Appl  
38 58 12.4 378 7 US-11-087-099-8872 Sequence 8872, Ap  
39 58 12.4 378 7 US-11-087-099-11183 Sequence 11183, A  
40 58 12.4 391 7 US-11-207-626A-32 Sequence 32, Appl  
41 58 12.4 451 7 US-11-124-620-5 Sequence 5, Appl  
42 58 12.4 451 7 US-11-124-620-7 Sequence 7, Appl  
43 58 12.4 464 7 US-11-087-099-7240 Sequence 7240, Ap  
44 58 12.4 2204 7 US-11-052-554A-134 Sequence 134, App  
45 57.5 12.3 108 7 US-11-097-812-39 Sequence 39, Appl

#### ALIGNMENTS

RESULT 1  
US-10-517-939-328  
; Sequence 328, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Blum, David  
; APPLICANT: Esteghlalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 328  
; LENGTH: 499  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(29)  
US-10-517-939-328

Query Match 34.8%; Score 163; DB 6; Length 499;  
Best Local Similarity 35.3%; Pred. No. 9.3e-12;  
Matches 30; Conservative 21; Mismatches 30; Indels 4; Gaps 2;  
QY 5 VKVQYKNDSPGDNQIKPGLQVNTGSSSDVISTVTYVWF-TRDGGSTLVYNCDAW 63  
DB 354 LSVQYRTGDSVNSNQIRPQIHVKNNSTLVNLKNTVRYWYNTKNGQN---PDCDYAK 410  
QY 64 MCGGNIRASFGSVNPATPTADTYLQ 88  
DB 411 ICGSVNTHKFTVLQKPKVGADAYLE 435  
RESULT 2

Query Match 24.1%; Score 113; DB 7; Length 382;  
Best Local Similarity 30.9%; Pred. No. 6.2e-06;  
Matches 29; Conservative 21; Mismatches 32; Indels 12; Gaps 4;

; GENERAL INFORMATION:  
 ; APPLICANT: Steer, Brian  
 ; APPLICANT: Callen, Walter  
 ; APPLICANT: Healey, Shaun  
 ; APPLICANT: Hazlewood, Geoff  
 ; APPLICANT: Blum, Di  
 ; APPLICANT: Blum, David  
 ; APPLICANT: Esteghalian, Al  
 ; TITLE OF INVENTION: XYLANASE  
 ; TITLE OF INVENTION: AND MET

; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 344  
; LENGTH: 738  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample.  
US-10-517-939-344

Query Match 23.0%; Score 108; DB 6; Length 738;  
Best Local Similarity 32.1%; Pred. No. 5.2e-05;  
Matches 27; Conservative 18; Mismatches 31; Indels 8; Gaps 4;

QY 7 VOYKNDAPGDN--QIKPGLQLVNTGSSVDLSTVTYRYWFTRDGSGSTLVYNCDAAM 64  
DB 591 LQYR---SADGNNYQMKPQTIKNAGKVPILSELTIRYFTPE-STQPVDTRIDNAQF 646

QY 65 GCGNIRASFGSVNPTATPTADTYLQ 88  
DB 647 GAHVQTTV--VPPSDAAAHAYVE 668

## RESULT 6

US-11-170-653-51  
; Sequence 51, Application US/11170653  
; Publication No. US20050271769A1  
; GENERAL INFORMATION:

; APPLICANT: Danisco A/S  
; APPLICANT: Sibbesen, Ole  
; APPLICANT: Sorensen, Jens  
; TITLE OF INVENTION: Xylanase Variants Having Altered Sensitivity to Xylanase Inhibitor  
; FILE REFERENCE: 674509-2046  
; CURRENT APPLICATION NUMBER: US/11/170,653  
; CURRENT FILING DATE: 2005-06-23  
; PRIOR APPLICATION NUMBER: US/10/237,386  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: PCT/IB01/00426  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: GB 0005585.5  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: GB 0015751.1  
; PRIOR FILING DATE: 2000-06-27  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 51  
; LENGTH: 210  
; TYPE: PRT  
; ORGANISM: A. tubigenis  
US-11-170-653-51

Query Match 15.4%; Score 72; DB 7; Length 210;  
Best Local Similarity 31.5%; Pred. No. 0.23;  
Matches 23; Conservative 10; Mismatches 22; Indels 18; Gaps 4;

QY 29 NTGSSVDLSTVTYRYWFTRDG-----GSSLVYNCDAAMCGNIRASFG 74  
DB 40 NLGDTYDESAGTFSMYW--BDGVSDFFVWGLGTGSSSTITYSAEYASGSASLYAVYG 97

QY 75 SVNPTATPTADTYL 87  
DB 98 WVN--YPOAEYI 108

## RESULT 7

US-11-052-554A-281

; Sequence 281, Application US/11052554A  
; Publication No. US20050288666A1  
; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 281  
; LENGTH: 5291  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-281

Query Match 14.9%; Score 70; DB 7; Length 5291;  
Best Local Similarity 33.3%; Pred. No. 18;  
Matches 29; Conservative 15; Mismatches 19; Indels 24; Gaps 7;

QY 17 GD---NQIKPGLQLVNTGSS--VDL-STVTYRYWFTRDG---SSTLVYNCDA----- 61  
DB 848 GDDVINAVERGETLVVSGSTGVEAGQTVTTF-----GGKNYTTTVEANGSWTVNVPPA 902

QY 62 ---AAM--GCGNIRASFGSVNPTATPTAD 84  
DB 903 DLAALPDGAGNVQASVSNINGNSAQAD 929

## RESULT 8

US-11-072-512-2825  
; Sequence 2825, Application US/11072512  
; Publication No. US20060029945A1  
; GENERAL INFORMATION:

; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKI, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: 084335-0191  
; CURRENT APPLICATION NUMBER: US/11/072,512  
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/350,978  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: JP 2001-379298  
; PRIOR FILING DATE: 2001-11-05  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2825  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-072-512-2825

```
Query Match          14.6%; Score 68.5; DB 7; Length 500;
Best Local Similarity 33.3%; Pred. No. 1.7;
Matches 19; Conservative 9; Mismatches 18; Indels 11; Gaps 3;

Qy 31 GSSVDLSTVTVRYWFTRDG-----GSS--TLVYNCD----WAAMGCNIRASFGSV 76
Db 414 GTPETDIDSSCRYLKADGTQCPSGSGSYTVITCTCFISAYGARGSANIKVTFISV 470

RESULT 9
US-10-517-939-330
; Sequence 330, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 330
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample.
; NAME/KEY: SIGNAL
; LOCATION: (1)...(35)
US-10-517-939-330

Query Match          14.5%; Score 68; DB 6; Length 755;
Best Local Similarity 30.1%; Pred. No. 3.1;
Matches 31; Conservative 8; Mismatches 40; Indels 24; Gaps 5;

Qy 2 SGGVKVOYK-NNDAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRD----- 49
Db 651 SGACTVAYAITNDWGS---FTANVTLTNTGSA--LNGWTLAYAFPQNOTISNAWNGTA 705

Qy 50 --GSSSTLVYNCDAWAMCGNIRASFG-----SVNPATPTADT 85
Db 706 VQSGSSVSVTNAGMNSGLPFPNVSASFGFQASYSNGNSVPSAFT 748

RESULT 10
US-11-207-626A-26
; Sequence 26, Application US/11207626A
; Publication No. US20060014276A1
; GENERAL INFORMATION:
; APPLICANT: Havenga, Menzo
; APPLICANT: Vogels, Ronald
; APPLICANT: Bout, Abraham
; TITLE OF INVENTION: CHIMERIC ADENOVIRUSES
; FILE REFERENCE: 2578-4123.2US
; CURRENT APPLICATION NUMBER: US/11/207,626A
; CURRENT FILING DATE: 2005-08-18
; PRIOR APPLICATION NUMBER: EP 98202297.2
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
```

```
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Human Adenovirus 32 Fiber Protein
US-11-207-626A-26

Query Match          13.8%; Score 64.5; DB 7; Length 389;
Best Local Similarity 32.0%; Pred. No. 3.8;
Matches 24; Conservative 10; Mismatches 32; Indels 9; Gaps 3;

Qy 11 NNDAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCDAWAMCGNIR 70
Db 251 NNNTNPTDKKITVKL-LFNEKGVLMDSLSLKEYWNYRNDNSTSQAY--DNAVPFMPNLIK 307

Qy 71 ASFGSVNPATPTADT 85
Db 308 AY-----PKPTTDT 316

RESULT 11
US-11-207-626A-20
; Sequence 20, Application US/11207626A
; Publication No. US20060014276A1
; GENERAL INFORMATION:
; APPLICANT: Havenga, Menzo
; APPLICANT: Vogels, Ronald
; APPLICANT: Bout, Abraham
; TITLE OF INVENTION: CHIMERIC ADENOVIRUSES
; FILE REFERENCE: 2578-4123.2US
; CURRENT APPLICATION NUMBER: US/11/207,626A
; CURRENT FILING DATE: 2005-08-18
; PRIOR APPLICATION NUMBER: EP 98202297.2
; PRIOR FILING DATE: 1998-07-08
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Human Adenovirus 24 Fiber Protein
US-11-207-626A-20

Query Match          13.2%; Score 62; DB 7; Length 389;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches 25; Conservative 9; Mismatches 33; Indels 8; Gaps 3;

Qy 11 NNDAPGDNOIKPGLQVLNTGSSVDLSTVTVRYWFTRDGSSSTLVYNCDAWAMCGNIR 70
Db 250 NNNTNPTDKKITVKL-LFNEKGVLMDSLSLKEYWNYRNDNSTSQAY--DNAVPFMPNLIK 307

Qy 71 ASFGSVNPATPTADT 85
Db 308 AY-----PKPTTDT 316

RESULT 12
US-10-517-939-14
; Sequence 14, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; CURRENT FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
```

```
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-517-939-14

Query Match      13.0%; Score 61; DB 6; Length 350;
Best Local Similarity 26.7%; Pred. No. 8.7;
Matches 16; Conservative 10; Mismatches 26; Indels 8; Gaps 1;

Qy      10 KNDSPAGDNGQIKPGLQVNTGSSVDLSVTIVRYVFTRDGSGSSTLVYNCOW 61
Db      40 RHNSLTADNQMPESVLDRTATLAKGDLHLAAVDFTRVDALMYFARDNGIPMYRHTLAW 99

RESULT 13
US-11-102-476-46
; Sequence 46, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U.M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102.476
; CURRENT FILING DATE: 2005-04-08
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1889
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-46

Query Match      13.0%; Score 61; DB 7; Length 1889;
Best Local Similarity 26.2%; Pred. No. 62;
Matches 21; Conservative 14; Mismatches 27; Indels 18; Gaps 4;

Qy      19 NQIKPGL-----QLVNTG-----SSVDLSVTIVRYVFTRDGSGSSTLVY-----NCDWA 62
Db      498 NPVNPSLAKGLTKFTATGIYSDNSNKDITSVTF--WFSSDSSIATISNAQKQGNAYGA 555

Qy      63 AMCGGNIRASFGSVNPAIPT 82
Db      556 ATGTTDIKATFGKVSPPVST 575

RESULT 14
US-11-098-686-10278
; Sequence 10278, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098.686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10278
; LENGTH: 806
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10278

Query Match      12.9%; Score 60.5; DB 7; Length 806;
Best Local Similarity 33.3%; Pred. No. 26;
Matches 20; Conservative 5; Mismatches 20; Indels 15; Gaps 4;

Qy      10 KNDSPAGDNGQIKPGLQVNTGSSVDLSVTIVRYVFT--TRDQ--GSSSTLVY--NCDWAA 63
Db      286 KNNIVTPGKDPVYPSSQLVRNHQ-----RYWVRGMADGFIGNSTWCIISNLDYVS 336

RESULT 15
US-11-087-099-7483
; Sequence 7483, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087.099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 7483
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Pseudomonas syringae pv. syringae B728a
US-11-087-099-7483

Query Match      12.9%; Score 60.5; DB 7; Length 983;
Best Local Similarity 33.3%; Pred. No. 33;
Matches 15; Conservative 9; Mismatches 10; Indels 11; Gaps 2;

Qy      2 SGGVKVQYKNDSDAPGD--NQIKPGLQ-----VNTGSSSV 35
Db      522 NAGVKAIFYFSNTSFGDPLVTRVEFGVNLNWTGSGVNTNAGSTAV 566

Search completed: March 2, 2006, 14:28:14
Job time : 1.78442 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:06:58 ; Search time 31.7439 Seconds  
(without alignments)  
4927.737 Million cell updates/sec

Title: US-09-917-376-5  
Perfect score: 469  
Sequence: 1 VSGGVKQVQKNDSPAGDQ.....TRAFSGVNPATPATDYIQ 88

Scoring table:  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O=/abs/ABSSWEB spool/US0917376/runat\_02032006\_091454\_8101/app\_query.fasta\_1  
-DB=Issued\_Patents\_NA -QFMT=fastcap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs06h -USER=US0917376 @CGN 1.1.278 @runat\_02032006\_091454\_8101  
-NCPU=6 -ICPU=3 -NO WMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/1 COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq.\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq.\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID                | Description       |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1          | 195.5 | 41.7        | 2029   | US-09-136-574A-46 | Sequence 46, Appl |
| 2          | 194.5 | 41.5        | 6416   | US-09-136-574A-2  | Sequence 2, Appl  |
| 3          | 192.5 | 41.0        | 11707  | US-09-136-574A-1  | Sequence 1, Appl  |
| 4          | 182.5 | 38.9        | 2977   | US-07-862-588B-1  | Sequence 1, Appl  |
| 5          | 165   | 35.2        | 1438   | US-09-339-159B-3  | Sequence 3, Appl  |
| 6          | 165   | 35.2        | 1482   | US-09-198-956-9   | Sequence 9, Appl  |
| 7          | 165   | 35.2        | 1482   | US-09-198-955A-11 | Sequence 11, Appl |
| 8          | 165   | 35.2        | 1482   | US-09-694-531-11  | Sequence 11, Appl |
| 9          | 165   | 35.2        | 1482   | US-09-670-141-9   | Sequence 9, Appl  |

|    |       |      |        |   |                     |                    |
|----|-------|------|--------|---|---------------------|--------------------|
| 10 | 165   | 35.2 | 1482   | 3 | US-10-072-152-11    | Sequence 11, Appl  |
| 11 | 155.5 | 33.2 | 4059   | 3 | US-09-784-554B-1    | Sequence 1, Appl   |
| 12 | 150.5 | 32.1 | 4056   | 3 | US-09-784-554B-3    | Sequence 3, Appl   |
| 13 | 120.5 | 25.7 | 1624   | 2 | US-07-862-588B-6    | Sequence 6, Appl   |
| 14 | 120.5 | 25.7 | 1775   | 2 | US-07-862-588B-5    | Sequence 5, Appl   |
| 15 | 114   | 24.3 | 486    | 2 | US-08-048-164A-1    | Sequence 1, Appl   |
| 16 | 114   | 24.3 | 486    | 2 | US-08-048-164A-3    | Sequence 3, Appl   |
| 17 | 114   | 24.3 | 486    | 2 | US-08-460-462-1     | Sequence 1, Appl   |
| 18 | 114   | 24.3 | 486    | 2 | US-08-460-462-3     | Sequence 3, Appl   |
| 19 | 114   | 24.3 | 486    | 2 | US-08-460-457-1     | Sequence 1, Appl   |
| 20 | 114   | 24.3 | 486    | 2 | US-08-460-457-3     | Sequence 3, Appl   |
| 21 | 114   | 24.3 | 486    | 2 | US-08-460-458-1     | Sequence 1, Appl   |
| 22 | 114   | 24.3 | 486    | 2 | US-08-460-458-3     | Sequence 3, Appl   |
| 23 | 114   | 24.3 | 486    | 2 | US-08-460-455-1     | Sequence 1, Appl   |
| 24 | 114   | 24.3 | 486    | 2 | US-08-460-455-3     | Sequence 3, Appl   |
| 25 | 114   | 24.3 | 486    | 2 | US-08-330-394A-1    | Sequence 1, Appl   |
| 26 | 114   | 24.3 | 486    | 2 | US-08-330-394A-3    | Sequence 3, Appl   |
| 27 | 114   | 24.3 | 499    | 3 | US-09-006-636-5     | Sequence 5, Appl   |
| 28 | 114   | 24.3 | 499    | 3 | US-09-006-632-5     | Sequence 5, Appl   |
| 29 | 114   | 24.3 | 499    | 3 | US-09-325-274-5     | Sequence 5, Appl   |
| 30 | 113   | 24.1 | 1146   | 3 | US-09-277-716-21    | Sequence 21, Appl  |
| 31 | 113   | 24.1 | 1146   | 3 | US-09-609-161B-21   | Sequence 21, Appl  |
| 32 | 78.5  | 16.7 | 764    | 3 | US-09-533-559-5404  | Sequence 5404, Ap  |
| 33 | 77    | 16.4 | 8107   | 3 | US-09-335-586-3     | Sequence 3, Appl   |
| 34 | 76    | 16.2 | 136058 | 3 | US-09-949-016-12565 | Sequence 12565, A  |
| 35 | 76    | 16.2 | 136480 | 3 | US-09-949-016-17064 | Sequence 17064, A  |
| 36 | 75.5  | 16.1 | 5714   | 3 | US-09-620-312D-393  | Sequence 393, App  |
| 37 | 72.5  | 15.5 | 819    | 3 | US-09-902-540-3752  | Sequence 3752, Ap  |
| 38 | 72.5  | 15.5 | 17897  | 3 | US-09-902-540-1182  | Sequence 1182, Ap  |
| 39 | 72.5  | 15.5 | 20284  | 3 | US-09-526-193A-21   | Sequence 21, Appl  |
| 40 | 70    | 14.9 | 10892  | 3 | US-09-902-540-962   | Sequence 962, Appl |
| 41 | 70    | 14.9 | 25165  | 3 | US-09-453-702B-39   | Sequence 39, Appl  |
| 42 | 70    | 14.9 | 25165  | 3 | US-10-114-170-39    | Sequence 39, Appl  |
| 43 | 70    | 14.9 | 246444 | 3 | US-09-949-016-13113 | Sequence 13113, A  |
| 44 | 69.5  | 14.8 | 103987 | 3 | US-09-949-016-12513 | Sequence 12513, A  |
| 45 | 69.5  | 14.8 | 103988 | 3 | US-09-949-016-17050 | Sequence 17050, A  |

ALIGNMENTS

RESULT 1  
US-09-136-574A-46  
; Sequence 46, Application US/09136574A  
; Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson

STREET: Spring House Corporate Center, P.O. Box 457

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

```

; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2029 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46
Alignment Scores:
Pred. No.: 5,09e-17 Length: 2029
Score: 195.50 Matches: 38
Percent Similarity: 60.2% Conservative: 15
Best Local Similarity: 43.2% Mismatches: 34
Query Match: 41.7% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-46 (1-2029)
QY 1 ValSerGlyValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1 ATGGAGAGTGTGTGAAGGTACTGTACAGAAACATGAGCAAGTGGAGCACAGTTCT 60
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 61 ATAAGCGCGTGTAAAGTAGTAGTGAATGAGGAGCAGCAGTGTGTATCTTAGCAGGGTT 120
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 121 AAGATAAGATAGTGTACACAGTGTGATGTCACAAAGCCACAGAGTGGCGTA--TGTGAC 177
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 178 TGGCCACAGATAGGCGCAAGCAATGTGACATCAATTTTGTGAAGCTTAGCAGCGGAGTG 237
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 238 AGTGGAGCGGATTAATCTGGAG 261

RESULT 2
US-09-136-574A-2
; Sequence 2, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Center, P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,574A
; FILING DATE: 19-Aug-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/932,571
; FILING DATE: September 19, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: 1997US001/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-136-574A-2
Alignment Scores:
Pred. No.: 3,72e-16 Length: 6416
Score: 194.50 Matches: 38
Percent Similarity: 61.2% Conservative: 14
Best Local Similarity: 44.7% Mismatches: 32
Query Match: 41.5% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-2 (1-6416)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 2665 GGTGTGAAGTACTGTACAGAACATGAGCAAGTGGAGCAGCAGGTTCTATAGGCGG 2724
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 2725 TGGTTTAAAGATAGTAGTGAATGAGGAGCAGCAGCAGTGTGTATCTTAGCAGCGTAAAGATAAGA 2784
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
Db 2785 TACTGTACACAGTGTGATGGTGACAAAGCCACAGAGTGGCGTA---TGTGACTGGGCACAG 2841
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 2842 ATAGGGCAAGCATGTGACATTCATTTTGTGAAGCTTAGCAGCGGAGTGGAGCGG 2901
QY 84 AspThrTyrLeuGln 88
Db 2902 GATTATTACCTGGAG 2916

RESULT 3
US-09-136-574A-1
; Sequence 1, Application US/09136574A
; Patent No. 6294366
; GENERAL INFORMATION:
; APPLICANT: Farrington, Graham K.
; Anderson, Paige
; Gibbs, Moreland
; Bergquist, Peter
; Daniels, Roy
; Morgan, Hugh W.
; Williams, Diane P.
; TITLE OF INVENTION: Compositions and Methods for
; Treating Cellulose Containing Fabrics Using Truncated
; Cellulase Enzyme Compositions
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:

```



ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997

## ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 11707 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-136-574A-1

Alignment Scores:  
Pred. No.: 1,698-15 Length: 11707  
Score: 192.50 Matches: 37  
Percent Similarity: 62.4% Conservative: 16  
Best Local Similarity: 43.5% Mismatches: 31  
Query Match: 41.0% Indels: 1  
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-136-574A-1 (1-11707)

QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23  
DB 4038 GGTGGAGGTACTATACAGACACATGAGCAAGTGCAGCACAAGTTCTATAGCGCG 4097  
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43  
DB 4098 TGGTTTAAGATAGTAGTGAATGGAGGCGAGCAGTGTGATCTTAGCAGGCTTAAGATAAGA 4157  
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63  
DB 4158 TACTGGTACACAGTGGATGGTGTGACAAAGCCACAGAGTGGCGTA---TGTACTGGGCACAG 4214  
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83  
DB 4215 ATAGGGGCAACCAATGTGACATTCATTTTGTGAAGCTGAGCAGCGGAGTGAGTGAGCG 4274  
QY 84 AspThrTyrLeuGln 88  
DB 4275 GATTATTACTTGGAG 4289

## RESULT 4

US-07-862-588B-1  
; Sequence 1, Application US/07862588B  
; Patent No. 5916796  
; GENERAL INFORMATION:  
; APPLICANT: Joergensen, Per Linea  
; APPLICANT: Sch lein, Martin  
; APPLICANT: Hansen, Christian  
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity

NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.  
STREET: 405 Lexington Avenue, 62nd floor  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10017  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/862,588B  
FILING DATE: 19920727  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 164/90  
FILING DATE: 19-JAN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00013  
FILING DATE: 18-JAN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Zelson, Steve T. / Lambiris, Elias J.  
REGISTRATION NUMBER: 30,335 / 33,728  
REFERENCE/DOCKET NUMBER: 3425.204-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2977 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Bacillus lautus  
STRAIN: NCIMB 40250  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 677..2776  
OTHER INFORMATION:  
US-07-862-588B-1

Alignment Scores:  
Pred. No.: 5,76e-15 Length: 2977  
Score: 182.50 Matches: 36  
Percent Similarity: 66.3% Conservative: 19  
Best Local Similarity: 43.4% Mismatches: 25  
Query Match: 38.9% Indels: 3  
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-1 (1-2977)

QY 6 LysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysProGlyLeu 25  
DB 144 CGACTGCATACAGCGCGCGCCGATACAAATGACGACGACCAACAGATCAAGCGCTCCTTC 203  
QY 26 GlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyrTrp 45  
DB 204 AACATCAAAACACACCGTACTTCGGCTGTGATTTAAGCACGCTCAAAATCCGCTACTAC 263  
QY 46 PheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrpAlaAlaMetGly 65  
DB 264 TTCACCAAGATGGTTCTCGCGCGGTGAACGGCTGG---ATCGACTGGGCGCAGCTCGGC 320  
QY 66 CysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAlaAspThr 85  
DB 321 GGCAGCAACATTCAGATCTCGTTTGGC-----AACCATACTGGCAGCAATTCGGATACG 374  
QY 86 TyrLeuGln 88

```
|||||: 375 TACGTGGAG 383
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-198-956-9

Alignment Scores:
Pred. No.: 5,78e-13 Length: 1482
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-956-9 (1-1482)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 991 GTATCAGGCAATTGGAAGTTGAATCTACACAGCAATCCTTCAGATACTACTACTCA 1050
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1051 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAAATC 1110
QY 41 ThrValArgTyrTrpPheThrArgGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1111 ACATTGAGATATTATTATACAGTAGACGAGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1168 CATGCTGCAATAATCGGCAGTAACGCGAGTACACGGAATTACTTCAATGTAAAGGA 1227
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1228 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGAGACACCTACCTTGAA 1278

RESULT 7
US-09-198-955A-11
; Sequence 11, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6187580el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/198,955A
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482

Db 375 TACGTGGAG 383
RESULT 5
US-09-339-159B-3
; Sequence 3, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-09-339-159B-3

Alignment Scores:
Pred. No.: 5,53e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-339-159B-3 (1-1438)
QY 1 ValSerGlyGlyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTGGAAGTTGAATCTACACAGCAATCCTTCAGATACTACTACTCA 999
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAGGTTACTAATACCGAAGCAGTGCAATTGATTGTCCAAATC 1059
QY 41 ThrValArgTyrTrpPheThrArgGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1060 ACATTGAGATATTATTATACAGTAGACGAGACAGAAAGATCAGACCTTCTGG---TGTGAC 1116
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTAACGCGAGTACACGGAATTACTTCAATGTAAAGGA 1176
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAAAAATGAGTTCTCTCAACAAATAACGAGACACCTACCTTGAA 1227

RESULT 6
US-09-198-956-9
; Sequence 9, Application US/09198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
```

TYPE: DNA  
ORGANISM: Clostridium thermocellum  
US-09-198-955A-11

Alignment Scores:  
Pred. No.: 5.78e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservative: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-198-955A-11 (1-1482)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGCAATTGAAGTTGAATTTCTACACAGCAATCTTCACTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCTCAAGTTCAAGTTACTAATACCGAAGCAGTGCATTTGTTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTACCGCAGCTACACGGAATTTACTTCAAAATGTAAGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTAAAATGAGTTCTCTCAACAATAACGACAGACCTTCTTGAA 1278

## RESULT 8

US-09-694-531-11  
; Sequence 11, Application US/09694531  
; Patent No. 6368843  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. 6368843el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/09/694,531  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 09/198,955  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum

US-09-694-531-11

TYPE: DNA  
ORGANISM: Clostridium thermocellum

## Alignment Scores:

Pred. No.: 5.78e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservative: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-694-531-11 (1-1482)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGCAATTGAAGTTGAATTTCTACACAGCAATCTTCACTACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCTCAAGTTCAAGTTACTAATACCGAAGCAGTGCATTTGTTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTAGACGACAGAAAGATCAGACCTTCTGG---TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTACCGCAGCTACACGGAATTTACTTCAAAATGTAAGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTAAAATGAGTTCTCTCAACAATAACGACAGACCTTCTTGAA 1278

## RESULT 9

US-09-670-141-9  
; Sequence 9, Application US/09670141  
; Patent No. 6429000  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels Erik K.  
; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Schnorr, Kirk  
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus  
; TITLE OF INVENTION: Licheniformis  
; FILE REFERENCE: 5377.200-US  
; CURRENT APPLICATION NUMBER: US/09/670,141  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: 09/198,956  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis

US-09-670-141-9

## Alignment Scores:

Pred. No.: 5.78e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservative: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 3 Gaps: 2

US-09-917-376-5 (1-88) x US-09-670-141-9 (1-1482)

QY 1 ValSerGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20  
Db 991 GTATCAGCAATTGAAGTTGAATTTCTACACAGCAATCTTCACTACTACTCA 1050



```
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784, 554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 4056
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-3

Alignment Scores:
Pred. No.: 2, 61e-10 Length: 4056
Score: 150.50 Matches: 27
Percent Similarity: 56.5% Conservative: 21
Best Local Similarity: 31.8% Mismatches: 36
Query Match: 32.1% Indels: 1
DB: 3 Gaps: 1

US-09-917-376-5 (1-88) x US-09-784-554B-3 (1-4056)
QY 4 GlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
Db 3595 GGATTGGTCTCCAGTATCGCACGGGATCAAAATGTGAACGACATCACTTGAACCG 3654
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
Db 3655 CATTTCCAAATTTAAATAAAGGTACATCTCCGTACCGATCAACGAGTTGAAATTCGC 3714
QY 44 TyrTrpPheThrArgAspGlyGlySerThrLeuValThrValThrValAla 63
Db 3715 TACTACTACGATCGCGGTGACCGTGACGAG---ACATTCACCTGCGACTATCGGTG 3771
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 3772 CTGACTCTCTGAACCTGAATGTTAAGTGGTTAAATGATAAAGCTGCAACCGTGCT 3831
QY 84 AspThrThrLeuGln 88
Db 3832 GATTATTATTGGAA 3846

RESULT 13
US-07-862-588B-6
; Sequence 6, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862.588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991

; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1624 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..1607
; OTHER INFORMATION:
US-07-862-588B-6

Alignment Scores:
Pred. No.: 1, 06e-06 Length: 1624
Score: 120.50 Matches: 28
Percent Similarity: 54.1% Conservative: 18
Best Local Similarity: 32.9% Mismatches: 36
Query Match: 25.7% Indels: 3
DB: 2 Gaps: 2

US-09-917-376-5 (1-88) x US-07-862-588B-6 (1-1624)
QY 2 SerGlyGlyVallyValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1323 ACGGGAACTTGTTCCTCAATAACAAAGTTGGGACACTAGCGCCACGATAACCAATG 1382
QY 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1383 AAGCTTCTCTTAACATCAAGAACACGGTACAAACCCCTGTTAACCTGAGCGGCTCAAG 1442
QY 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 1443 CTTNNNNNNNNNNNNAAAAAGAC---GGACCTGGCGGATATGAGTCTCGATCGACTGG 1499
QY 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1500 GCGCAATCGCGCAACGAATGTTCTGCTGGCATTG-----GCTAACTTTACCGGGAGT 1553
QY 82 ThrAlaAspThrTyr 86
Db 1554 AATACGGATACTTAC 1568

RESULT 14
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linnaa
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB 40250
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..(1625,1775)
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /partial
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
; OTHER INFORMATION: OTR)
US-07-862-588B-5

```

```

Alignment Scores:
Pred. No.: 1.2e-06 Length: 1775
Score: 120.50 Matches: 28
Percent Similarity: 54.1% Conservative: 18
Best Local Similarity: 32.9% Mismatches: 36
Query Match: 25.7% Indels: 3
DB: 2 Gaps: 2

```

US-09-917-376-5 (1-88) x US-07-862-588B-5 (1-1775)

```

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 1323 ACGGGGAACCTGTGTGTCATACAAAGTTCGCGACACTGCGCCACGATAACAAATG 1382

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 1383 AAGCTTCTTTTACATCAAGAACACGTCACACCCCTGTACCTGAGCGGCTCAAG 1442

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerThrLeuValTyrAsnCysAspTrp 61
Db 1443 CTTNNNNNNNNNNNNAAAAAGAC--GGACCTGGGATATGAGTGTCTGATCGCTG 1499

Qy 62 AlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 1500 GCGCAATCGGCCGAACGAATGTTCTGTGTCGCAATTC-----GCTAACTTTACCGGGAGT 1553

Qy 82 ThrAlaAspThrTyr 86
Db 1554 AATACGGATACCTTAC 1568

RESULT 15
US-08-048-164A-1

```

```

; Sequence 1, Application US/08048164A
; Patent No. 5496934
; GENERAL INFORMATION:
; APPLICANT: Shoseyov, Oded
; APPLICANT: Shiegl, Itai
; APPLICANT: Goldstein, Marc A.
; APPLICANT: Doi, Roy H.
; TITLE OF INVENTION: CELLULOSE BINDING DOMAIN
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSER: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/048,164A
; FILING DATE: 14-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7809-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..486
US-08-048-164A-1

Alignment Scores:
Pred. No.: 1.49e-06 Length: 486
Score: 114.00 Matches: 29
Percent Similarity: 52.6% Conservative: 22
Best Local Similarity: 29.9% Mismatches: 34
Query Match: 24.3% Indels: 12
DB: 2 Gaps: 4

US-09-917-376-5 (1-88) x US-08-048-164A-1 (1-486)

Qy 2 SerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIle 21
Db 7 ACATCATCAATGTCAGTTGAATTTTCAACTCTAACAAATTCACCAAACTCAAACTCAATT 66

Qy 22 LysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThr 41
Db 67 ACACCAATAATCAAAATTAATTAACATCTGACAGTGATTAAATTTAAATGACGTAAAA 126

Qy 42 ValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrp 61
Db 127 GTTAGATATATTACACAAAGTGTGTCACACAAAGGACAACTTCTCGG---TGTGACCAT 183

Qy 62 AlaAlaMetGlyCysGlyAsn-----IleArgAlaSerPhe 73
Db 184 GCTGTGTCATTATTAGGAATAGCTATGTGTATACACTAGCAAAAGTACAGCAAACTTC 243

Qy 74 -----GlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88

```

Db 244 GTTAAGAAACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 291

Search completed: March 2, 2006, 13:50:35  
Job time : 36.7439 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:19:26 ; Search time 165.153 Seconds  
(without alignments)  
4406.259 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQVKNDSAPGNQ.....IRASFGVNPATPTADTYLQ 88

Scoring table:

|                           |                           |
|---------------------------|---------------------------|
| BLOSUM62                  | Xgapop 10.0 , Xgapext 0.5 |
| Ygapop 10.0 , Ygapext 0.5 |                           |
| Fgapop 6.0 , Fgapext 7.0  |                           |
| Delop 6.0 , Delext 7.0    |                           |

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DB=xlsp  
-Q=/abs/ABSWEB\_spool/US09917376/runat\_02032006\_091500\_8223/app\_query.fasta.1  
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs06p  
-USER=US09917376 @CGN 1 1 3095 @runat\_02032006\_091500\_8223 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:

|     |   |
|-----|---|
| 1:  | /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*  |
| 2:  | /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*  |
| 3:  | /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq* |
| 4:  | /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq* |
| 5:  | /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq* |
| 6:  | /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq* |
| 7:  | /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq* |
| 8:  | /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq* |
| 9:  | /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq* |
| 10: | /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq*  |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length  | DB ID | Description       |
|------------|-------|-------------|---------|-------|-------------------|
| 1          | 469   | 100.0       | 2289    | 3     | US-09-917-378-2   |
| 2          | 466   | 99.4        | 3687    | 3     | US-09-917-384-2   |
| 3          | 466   | 99.4        | 3687    | 3     | US-09-917-383-2   |
| 4          | 462   | 98.5        | 2869    | 3     | US-09-917-376-2   |
| 5          | 462   | 98.5        | 2869    | 5     | US-10-155-400-2   |
| 6          | 257   | 54.8        | 2223    | 6     | US-10-156-761-550 |
| 7          | 257   | 54.8        | 9025608 | 6     | US-10-156-761-1   |

|    |       |      |       |   |                      |
|----|-------|------|-------|---|----------------------|
| 8  | 190   | 40.5 | 1527  | 6 | US-10-369-493-46838  |
| 9  | 188   | 40.1 | 1434  | 8 | US-10-466-208-11     |
| 10 | 188   | 40.1 | 1491  | 8 | US-10-466-208-7      |
| 11 | 188   | 40.1 | 2510  | 7 | US-10-433-577-34     |
| 12 | 165   | 35.2 | 501   | 9 | US-10-933-404-1      |
| 13 | 165   | 35.2 | 1438  | 6 | US-10-372-054-3      |
| 14 | 165   | 35.2 | 1482  | 5 | US-10-072-152-11     |
| 15 | 165   | 35.2 | 1482  | 7 | US-10-855-433-11     |
| 16 | 157   | 33.5 | 1314  | 3 | US-09-974-300-684    |
| 17 | 155.5 | 33.2 | 4059  | 3 | US-09-784-554B-1     |
| 18 | 155.5 | 33.2 | 4059  | 3 | US-10-896-555-1      |
| 19 | 150.5 | 32.1 | 4056  | 3 | US-09-784-554B-3     |
| 20 | 150.5 | 32.1 | 4056  | 8 | US-10-896-555-3      |
| 21 | 131.5 | 28.0 | 3180  | 9 | US-10-886-393A-29    |
| 22 | 130.5 | 27.8 | 4452  | 7 | US-10-282-122A-15299 |
| 23 | 126.5 | 27.0 | 1303  | 9 | US-10-490-737-1      |
| 24 | 126.5 | 27.0 | 1747  | 9 | US-10-490-737-2      |
| 25 | 126.5 | 27.0 | 3489  | 9 | US-10-490-737-3      |
| 26 | 114   | 24.3 | 768   | 6 | US-10-261-446-5      |
| 27 | 114   | 24.3 | 768   | 7 | US-10-261-446B-5     |
| 28 | 113   | 24.1 | 652   | 8 | US-10-782-234-5      |
| 29 | 113   | 24.1 | 652   | 8 | US-10-748-055-9      |
| 30 | 113   | 24.1 | 1146  | 3 | US-09-808-898-21     |
| 31 | 78.5  | 16.7 | 764   | 8 | US-10-853-047-5404   |
| 32 | 78.5  | 16.7 | 3147  | 7 | US-10-282-122A-13716 |
| 33 | 76    | 16.2 | 757   | 7 | US-10-767-701-8547   |
| 34 | 76    | 16.2 | 96597 | 7 | US-10-052-482-226    |
| 35 | 75.5  | 16.1 | 1041  | 7 | US-10-260-238-742    |
| 36 | 75.5  | 16.1 | 1041  | 7 | US-10-437-963-76751  |
| 37 | 75.5  | 16.1 | 2091  | 7 | US-10-211-462-222    |
| 38 | 75.5  | 16.1 | 2091  | 8 | US-10-723-860-4038   |
| 39 | 75.5  | 16.1 | 2091  | 9 | US-10-756-149-4006   |
| 40 | 75.5  | 16.1 | 3396  | 7 | US-10-398-458-1      |
| 41 | 75.5  | 16.1 | 4041  | 3 | US-09-747-835A-21    |
| 42 | 75.5  | 16.1 | 4041  | 7 | US-10-312-312-21     |
| 43 | 75.5  | 16.1 | 4353  | 5 | US-10-225-567A-405   |
| 44 | 75.5  | 16.1 | 4755  | 9 | US-10-505-486-201    |
| 45 | 75.5  | 16.1 | 4914  | 3 | US-09-974-298-188    |

## ALIGNMENTS

### RESULT 1

US-09-917-378-2  
; Sequence 2, Application US/09917378  
; Publication NO. US20030119093A1  
; GENERAL INFORMATION:  
; APPLICANT: DING, SHI-YOU  
; APPLICANT: ADNEY, WILLIAM S.  
; APPLICANT: VINZANT, TODD B.  
; APPLICANT: DECKER, STEPHEN R.  
; APPLICANT: HIMMEL, MICHAEL E.  
; TITLE OF INVENTION: THERMAL TOLERANT MANNANASE FROM ACIDOTHERMUS CELLULOLYTICUS  
; FILE REFERENCE: 40197.7US01  
; CURRENT APPLICATION NUMBER: US/09/917,378  
; CURRENT FILING DATE: 2001-07-28  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2289  
; TYPE: DNA  
; ORGANISM: Acidothermus cellulolyticus  
US-09-917-378-2

|                        |          |
|------------------------|----------|
| Alignment Scores:      |          |
| Pred. No.:             | 6.94e-56 |
| Score:                 | 469.00   |
| Percent Similarity:    | 100.0%   |
| Best Local Similarity: | 100.0%   |
| Query Match:           | 100.0%   |
| DB:                    | 3        |
| Length:                | 2289     |
| Matches:               | 88       |
| Conservative:          | 0        |
| Mismatches:            | 0        |
| Indels:                | 0        |
| Gaps:                  | 0        |



US-09-917-376-5 (1-88) x US-09-917-378-2 (1-2289)

```
QY 1 ValSerGlyValValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1363 GTGTCGGGTGGGTCAAGTGCAGTACAAACAATGATTCCGGCCGGGTGATAACACG 1422
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1423 ATCAAAACCGGCTCTCCAGTTCGTAATACGGGGTGTCTCGTGGTGGATTTCGACGGTG 1482
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1483 ACGGTGGGTACTGTTCCACCGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1542
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1543 TGGCGCGCATGGGTGTGGGAATATCCGCGCTCGTTGGCTGGTGAACCGCGGACG 1602
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1603 CCGACGGCGGACACCTACTCTGCAG 1626
```

## RESULT 2

US-09-917-384-2

; Sequence 2, Application US/09917384

; Publication No. US20030096342A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917.384

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 3687

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of

; OTHER INFORMATION: Guxa

US-09-917-384-2

## Alignment Scores:

```
Pred. No.: 3.5e-55 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 98.9% Mismatches: 0
Query Match: 99.4% Indels: 0
DB: 3 Gaps: 0
```

US-09-917-376-5 (1-88) x US-09-917-384-2 (1-3687)

```
QY 1 ValSerGlyValValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTCAAGTGCAGTACAAACAATGATTCCGGCCGGGTGATAACACG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGCTCTCCAGTTCGTAATACGGGGTGTCTCGTGGTGGATTTCGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGTACTGTTCCACCGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCATGGGTGTGGGAATATCCGCGCTCGTTGGCTGGTGAACCGCGGACG 1989
```

```
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACTCTGCAG 2013
```

## RESULT 3

US-09-917-383-2

; Sequence 2, Application US/09917383

; Publication No. US20030104522A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: DECKER, STEPHEN R.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS

; FILE REFERENCE: 40170.6US01

; CURRENT APPLICATION NUMBER: US/09/917.383

; CURRENT FILING DATE: 2001-07-28

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 3687

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Segment of

; OTHER INFORMATION: Guxa

US-09-917-383-2

## Alignment Scores:

```
Pred. No.: 3.5e-55 Length: 3687
Score: 466.00 Matches: 87
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 98.9% Mismatches: 0
Query Match: 99.4% Indels: 0
DB: 3 Gaps: 0
```

US-09-917-376-5 (1-88) x US-09-917-383-2 (1-3687)

```
QY 1 ValSerGlyValValValValGlnTyrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 1750 GTGTCGGGTGGGTCAAGTGCAGTACAAACAATGATTCCGGCCGGGTGATAACACG 1809
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1810 ATCAAAACCGGCTCTCCAGTTCGTAATACCGGGTGTCTCGTGGTGGATTTCGACGGTG 1869
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAsp 60
Db 1870 ACGGTGGGTACTGTTCCACCGGGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1929
QY 61 TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThr 80
Db 1930 TGGCGCGCATGGGTGTGGGAATATCCGCGCTCGTTGGCTGGTGAACCGCGGACG 1989
QY 81 ProThrAlaAspThrTyrLeuGln 88
Db 1990 CCGACGGCGGACACCTACTCTGCAG 2013
```

## RESULT 4

US-09-917-376-2

; Sequence 2, Application US/09917376

; Publication No. US20040038334A1

; GENERAL INFORMATION:

; APPLICANT: DING, SHI-YOU

; APPLICANT: ADNEY, WILLIAM S.

; APPLICANT: VINZANT, TODD B.

; APPLICANT: HIMMEL, MICHAEL E.

; TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS

; FILE REFERENCE: 40197.4US01





```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with additional his encoding sequences
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1431)
US-10-466-208-11
Alignment Scores:
Pred. No.: 4,53e-16 Length: 1434
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 8 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-11 (1-1434)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 991 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGAACAGCAACCAATCCGTCG 1050
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 1051 CAGCTTCAATATAAATAACCGCAATACCAAGTTGATTGATTAAAGATGTCACTGCCCGT 1110
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 1111 TACTGGTATACGCGAATAACAGGCAAAACGTT-----GACTGTGACTACGCGCAG 1164
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 1165 CTGTGATCGGCAATGTGACATACAAGTTTGTGACGTGTGCATAAACCAACCAAGGTTGCA 1224
QY 84 AspThrTyrLeuGln 88
DB 1225 GATACCTATCTGGAA 1239

RESULT 10
US-10-466-208-7
; Sequence 7, Application US/10466208
; Publication No. US20040180348A1
; GENERAL INFORMATION:
; APPLICANT: GENOFOCUS Co., Ltd.
; TITLE OF INVENTION: Method for Surface Display of Proteins on Genetic Carriers
; FILE REFERENCE: 3260-13
; CURRENT APPLICATION NUMBER: US/10/466,208
; CURRENT FILING DATE: 2003-07-14
; PRIOR APPLICATION NUMBER: KR2001-2156
; PRIOR FILING DATE: 2001-01-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 7
; LENGTH: 1491
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CMCase gene with mutated signal sequence to enhance its
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1488)
US-10-466-208-7
Alignment Scores:
Pred. No.: 4,77e-16 Length: 1491
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 8 Gaps: 1

US-09-917-376-5 (1-88) x US-10-466-208-7 (1-1491)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 991 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGAACAGCAACCAATCCGTCG 1050
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 1051 CAGCTTCAATATAAATAACCGCAATACCAAGTTGATTGATTAAAGATGTCACTGCCCGT 1110
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 1111 TACTGGTATACGCGAATAACAGGCAAAACGTT-----GACTGTGACTACGCGCAG 1164
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 1165 CTGTGATCGGCAATGTGACATACAAGTTTGTGACGTGTGCATAAACCAACCAAGGTTGCA 1224
QY 84 AspThrTyrLeuGln 88
DB 1225 GATACCTATCTGGAA 1239

RESULT 11
US-10-433-577-34
; Sequence 34, Application US/10433577
; Publication No. US20040171065A1
; GENERAL INFORMATION:
; APPLICANT: Genofocus Co., Ltd.
; TITLE OF INVENTION: Method for Expression of Proteins on Spore Surface
; FILE REFERENCE: PCT-Genofocus-1
; CURRENT APPLICATION NUMBER: US/10/433,577
; CURRENT FILING DATE: 2003-06-05
; PRIOR APPLICATION NUMBER: KR2000-74835
; PRIOR FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 34
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (461)..(2491)
US-10-433-577-34
Alignment Scores:
Pred. No.: 9,54e-16 Length: 2510
Score: 188.00 Matches: 34
Percent Similarity: 61.2% Conservative: 18
Best Local Similarity: 40.0% Mismatches: 31
Query Match: 40.1% Indels: 2
DB: 7 Gaps: 1

US-09-917-376-5 (1-88) x US-10-433-577-34 (1-2510)
QY 4 GlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnIleLysPro 23
DB 2051 GGTATTCTGTACATAACAGCAGCGGATGGAGTATGAACAGCAACCAATCCGTCG 2110
QY 24 GlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArg 43
DB 2111 CAGCTTCAATATAAATAACCGCAATACCAAGTTGATTGATTAAAGATGTCACTGCCCGT 2170
QY 44 TyrTrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAla 63
DB 2171 TACTGGTATACGCGAATAACAGGCAAAACGTT-----GACTGTGACTACGCGCAG 2224
QY 64 MetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
DB 2225 CTGTGATCGGCAATGTGACATACAAGTTTGTGACGTGTGCATAAACCAACCAAGGTTGCA 2284
QY 84 AspThrTyrLeuGln 88
DB 2285 GATACCTATCTGGAA 2299

```

```

RESULT 12
US-10-933-404-1
; Sequence 1, Application US/10933404
; Publication No. US20050118729A1
; GENERAL INFORMATION:
; APPLICANT: MORAG, Ely
; TITLE OF INVENTION: MICROARRAYS OF CELLULOSE BINDING CHIMERIC PROTEINS AND METHODS OF
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 85189-7100
; CURRENT APPLICATION NUMBER: US/10/933,404
; CURRENT FILING DATE: 2004-09-03
; PRIOR APPLICATION NUMBER: PCT/IL03/00177
; PRIOR FILING DATE: 2003-03-06
; PRIOR APPLICATION NUMBER: US 60/362,061
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Clostridium thermocellum
US-10-933-404-1

Alignment Scores:
Pred. No.: 2,13e-13 Length: 501
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 9 Gaps: 2

US-09-917-376-5 (1-88) x US-10-933-404-1 (1-501)

Qy 1 ValSerGlyValValGlnThrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 13 GTATCAGGCAATTTGAAGGTTGAATTTACACAGCAATCTTCAGATACCTAACTCA 72
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerThrLeuValThrVal 40
Db 73 ATCAATCCTCAGTTCAGGTTACTAATACCGAGCAGTGAATGATTGTTCCAAACTC 132
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrVal 60
Db 133 ACATTGAGATATTATTATATACAGTACGACAGAGATCAGACCTTCTGG---TGTGAC 189
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 190 CATGCTGCAATAATCGGCAGTACGACGAGCTACAAACGGAATTACTTCAAAATGTA 249
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 250 ACATTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTTACTTGA 300

RESULT 13
US-10-372-054-3
; Sequence 3, Application US/10372054
; Publication No. US20030203466A1
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: Novel Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/10/372,054
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US/09/339,159B
; PRIOR FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3

```

```

; LENGTH: 1438
; TYPE: DNA
; ORGANISM: Bacillus sp. 1633
US-10-372-054-3

Alignment Scores:
Pred. No.: 8,68e-13 Length: 1438
Score: 165.00 Matches: 35
Percent Similarity: 59.8% Conservative: 23
Best Local Similarity: 36.1% Mismatches: 29
Query Match: 35.2% Indels: 10
DB: 6 Gaps: 2

US-09-917-376-5 (1-88) x US-10-372-054-3 (1-1438)

Qy 1 ValSerGlyValValGlnThrLysAsnAspSerAlaProGlyAspAsnGln 20
Db 940 GTATCAGGCAATTTGAAGGTTGAATTTACACAGCAATCTTCAGATACCTAACTCA 999
Qy 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 1000 ATCAATCCTCAGTTCAGGTTACTAATACCGAGCAGTGAATGATTGTTCCAAACTC 1059
Qy 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrVal 60
Db 1060 ACATTGAGATATTATTATATACAGTACGACGACAGAGATCAGACCTTCTGG---TGTGAC 1116
Qy 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71
Db 1117 CATGCTGCAATAATCGGCAGTACGACGAGCTACAAACGGAATTACTTCAAAATGTA 1176
Qy 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88
Db 1177 ACATTGTAATAATGAGTTCTCTCAACAAATAACGACAGACCTTACTTGA 1227

RESULT 14
US-10-072-152-11
; Sequence 11, Application US/10072152
; Publication No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1el Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Clostridium thermocellum

```

## US-10-072-152-11

## Alignment Scores:

Pred. No.: 9,04e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservatives: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 2 Gaps: 5

US-09-917-376-5 (1-88) x US-10-072-152-11 (1-1482)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspGln 20  
Db 991 GTATCAGGCAATTTGAAGTTGAATTTCTACACAGCAATCCTTCAGATACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCTCAGTTCAAGGTTACTATACCGAAGCAGTGCATTTGTTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGACTACACGGAATTTACTTCAAAATGTAAGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTTCAACAAATAATACGCGACACCTACTCTTGA 1278

## RESULT 15

US-10-655-433-11  
; Sequence 11, Application US/10655433  
; Publication NO. US20040067572A1  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Lene N.  
; APPLICANT: Schulein, Martin  
; APPLICANT: Lange, Niels E.  
; APPLICANT: Bjoernvad, Mads E.  
; APPLICANT: Moller, Soren  
; APPLICANT: Glad, Sanne O. S.  
; APPLICANT: Kauppinen, Markus S.  
; APPLICANT: Schnorr, Kirk  
; APPLICANT: Kongsbak, Lars  
; TITLE OF INVENTION: No. US20040067572A1el Pectate Lyases  
; FILE REFERENCE: 5378.200-US  
; CURRENT APPLICATION NUMBER: US/10/655,433  
; CURRENT FILING DATE: 2003-09-04  
; PRIOR APPLICATION NUMBER: US/09/198,955A  
; PRIOR FILING DATE: 1998-11-24  
; PRIOR APPLICATION NUMBER: 1343/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 1344/97  
; PRIOR FILING DATE: 1997-11-24  
; PRIOR APPLICATION NUMBER: 60/067,249  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 60/067,240  
; PRIOR FILING DATE: 1997-12-02  
; PRIOR APPLICATION NUMBER: 09/073,684  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: 09/184,217  
; PRIOR FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID-NOS: 32  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Clostridium thermocellum  
US-10-655-433-11

## Alignment Scores:

Pred. No.: 9,04e-13 Length: 1482  
Score: 165.00 Matches: 35  
Percent Similarity: 59.8% Conservatives: 23  
Best Local Similarity: 36.1% Mismatches: 29  
Query Match: 35.2% Indels: 10  
DB: 2 Gaps: 5

US-09-917-376-5 (1-88) x US-10-655-433-11 (1-1482)

QY 1 ValSerGlyGlyValLysValGlnTyrLysAsnAspSerAlaProGlyAspGln 20  
Db 991 GTATCAGGCAATTTGAAGTTGAATTTCTACACAGCAATCCTTCAGATACTACTCA 1050  
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40  
Db 1051 ATCAATCTCTCAGTTCAAGGTTACTATACCGAAGCAGTGCATTTGTTCCAACTC 1110  
QY 41 ThrValArgTyrTrpPheThrArgAspGlyGlySerSerThrLeuValThrAsnCysAsp 60  
Db 1111 ACATTGAGATATTATTATACAGTACGACGACAGAAAGATCAGACCTTCTGG--TGTGAC 1167  
QY 61 TrpAlaAlaMet-----GlyCysGlyAsnIleArgAla 71  
Db 1168 CATGCTGCAATAATCGGCAGTAAACGCGACTACACGGAATTTACTTCAAAATGTAAGA 1227  
QY 72 SerPheGlySerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
Db 1228 ACATTGTGTAATAATGAGTTCTTCAACAAATAATACGCGACACCTACTCTTGA 1278

Search completed: March 2, 2006, 20:26:36  
Job time : 701.153 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 2, 2006, 12:24:59 ; Search time 94.4333 Seconds  
(without alignments)  
2043.020 Million cell updates/sec

Title: US-09-917-376-5

Perfect score: 469

Sequence: 1 VSGGVKQYKNDSAPGDNQ.....IRASFGSVNPTPTADTYLQ 88

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 7218512 seqs, 1096188070 residues

Total number of hits satisfying chosen parameters: 14437024

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame p2n.model -DEV=xlp  
-Q=/abs/ABSWEB/spool/US0917376/runat\_02032006\_091502\_8293/app\_query.fasta.1  
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -HOST=abs03p  
-USER=US0917376.ecgn\_1\_1067 @runat\_02032006\_091502\_8293 -NCPU=6 -ICPU=3  
-NO\_WMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.\*

1: /cgn2\_6/ptodata/2/pubpna/US08 NEW PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06 NEW PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07 NEW PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US11 NEW PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60 NEW PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | ID                | Description       |
|------------|-------|-------------|--------|-------------------|-------------------|
| 1          | 163   | 34.8        | 1500   | US-10-517-939-327 | Sequence 327, App |
| 2          | 157   | 33.5        | 2602   | US-10-510-386-229 | Sequence 229, App |
| 3          | 113   | 24.1        | 1146   | US-11-179-411-21  | Sequence 21, Appl |
| 4          | 113   | 24.1        | 1146   | US-11-175-766-21  | Sequence 21, Appl |

| 5  | 108  | 23.0 | 2217   | 8  | US-10-517-939-343     | Sequence 343, App  |
|----|------|------|--------|----|-----------------------|--------------------|
| 6  | 70   | 14.9 | 15876  | 12 | US-11-052-554A-660    | Sequence 660, App  |
| 7  | 69.5 | 14.8 | 617    | 6  | US-09-925-065A-353393 | Sequence 353393, A |
| 8  | 68.5 | 14.6 | 1084   | 8  | US-10-750-185-31182   | Sequence 31182, A  |
| 9  | 68.5 | 14.6 | 1084   | 8  | US-10-750-623-31182   | Sequence 31182, A  |
| 10 | 68.5 | 14.6 | 2231   | 9  | US-11-072-512-855     | Sequence 855, App  |
| 11 | 68.5 | 14.6 | 4952   | 12 | US-11-136-527-487     | Sequence 487, App  |
| 12 | 68   | 14.5 | 2268   | 8  | US-10-517-939-329     | Sequence 329, App  |
| 13 | 68   | 14.5 | 35344  | 8  | US-10-995-561-13307   | Sequence 13307, A  |
| 14 | 67.5 | 14.4 | 556    | 6  | US-09-925-065A-228751 | Sequence 228751, A |
| 15 | 67.5 | 14.4 | 2564   | 9  | US-11-072-512-279     | Sequence 279, App  |
| 16 | 67   | 14.3 | 523    | 6  | US-09-925-065A-39541  | Sequence 39541, A  |
| 17 | 67   | 14.3 | 17186  | 12 | US-11-121-086-105     | Sequence 105, App  |
| 18 | 66.5 | 14.2 | 600    | 8  | US-10-750-185-1174    | Sequence 1174, App |
| 19 | 66.5 | 14.2 | 600    | 8  | US-10-750-623-1174    | Sequence 1174, App |
| 20 | 66.5 | 14.2 | 1847   | 12 | US-11-000-688-1099    | Sequence 1099, App |
| 21 | 66.5 | 14.2 | 2565   | 8  | US-10-750-185-48986   | Sequence 48986, A  |
| 22 | 66.5 | 14.2 | 2565   | 8  | US-10-750-623-48986   | Sequence 48986, A  |
| 23 | 65.5 | 14.0 | 617    | 6  | US-09-925-065A-353394 | Sequence 353394, A |
| 24 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547306 | Sequence 547306, A |
| 25 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547307 | Sequence 547307, A |
| 26 | 65.5 | 14.0 | 1218   | 6  | US-09-925-065A-547308 | Sequence 547308, A |
| 27 | 64.5 | 13.8 | 600    | 6  | US-09-925-065A-336480 | Sequence 336480, A |
| 28 | 64.5 | 13.8 | 600    | 6  | US-09-925-065A-336481 | Sequence 336481, A |
| 29 | 64.5 | 13.8 | 601    | 6  | US-09-925-065A-729386 | Sequence 729386, A |
| 30 | 64.5 | 13.8 | 2370   | 6  | US-09-925-065A-91998  | Sequence 91998, A  |
| 31 | 64.5 | 13.8 | 2370   | 6  | US-09-925-065A-91999  | Sequence 91999, A  |
| 32 | 64.5 | 13.8 | 159138 | 8  | US-10-995-561-13230   | Sequence 13230, A  |
| 33 | 64   | 13.6 | 556    | 6  | US-09-925-065A-369647 | Sequence 369647, A |
| 34 | 64   | 13.6 | 585    | 6  | US-09-925-065A-385747 | Sequence 385747, A |
| 35 | 64   | 13.6 | 611    | 6  | US-09-925-065A-658978 | Sequence 658978, A |
| 36 | 64   | 13.6 | 659    | 6  | US-09-925-065A-87693  | Sequence 87693, A  |
| 37 | 64   | 13.6 | 659    | 6  | US-09-925-065A-87694  | Sequence 87694, A  |
| 38 | 64   | 13.6 | 1898   | 9  | US-11-072-512-1678    | Sequence 1678, App |
| 39 | 64   | 13.6 | 1990   | 12 | US-11-062-225-2       | Sequence 2, Appl   |
| 40 | 64   | 13.6 | 1990   | 12 | US-11-062-225-10      | Sequence 10, Appl  |
| 41 | 64   | 13.6 | 3074   | 12 | US-11-000-688-315     | Sequence 315, App  |
| 42 | 63.5 | 13.5 | 596    | 6  | US-09-925-065A-521812 | Sequence 521812, A |
| 43 | 63.5 | 13.5 | 597    | 6  | US-09-925-065A-258810 | Sequence 258810, A |
| 44 | 63.5 | 13.5 | 597    | 6  | US-09-925-065A-258811 | Sequence 258811, A |
| 45 | 63.5 | 13.5 | 606    | 6  | US-09-925-065A-484851 | Sequence 484851, A |

#### ALIGNMENTS

RESULT 1  
US-10-517-939-327  
; Sequence 327, Application US/10517939  
; Publication No. US20060003433A1  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Hazlewood, Geoff  
; APPLICANT: Wu, Di  
; APPLICANT: Blum, David  
; APPLICANT: Esteghalian, Alireza  
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
; FILE REFERENCE: 564462007901  
; CURRENT APPLICATION NUMBER: US/10/517,939  
; CURRENT FILING DATE: 2004-12-13  
; PRIOR APPLICATION NUMBER: PCT/US03/19153  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/389,299  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 327  
; LENGTH: 1500  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:

OTHER INFORMATION: Obtained from an environmental sample.

```
US-10-517-939-327
Alignment Scores:
Pred. No.: 8.87e-11 Length: 1500
Score: 163.00 Matches: 30
Percent Similarity: 60.0% Conservatives: 21
Best Local Similarity: 35.3% Mismatches: 30
Query Match: 34.8% Indels: 4
DB: 8 Gaps: 2

US-09-917-376-5 (1-88) x US-10-517-939-327 (1-1500)
Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24
Db 1060 CTATCTGTGTAATACAGACAGGGGATGTAACACCAACCAATCCGCTCAG 1119
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 1120 ATCATGTGAAACAAACACAGCAGCACCCTGTAATTTAAATAATGTAACCTGCTCAG 1179
Qy 45 TrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAspTrpAlaAla 63
Db 1180 TGGTATAACAGAAACAAAGGCCAAAC-----TTGACTGTGACTAGCGGAAG 1230
Qy 64 MetGlyCysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1231 ATCGGATGACCAATGTGACGACCAAGTTGTGACATTACAAACCTGTAAGAGGTGCA 1290
Qy 84 AspThrTyrLeuGln 88
Db 1291 GATGCCTATCTGGAA 1305

RESULT 2
US-10-510-386-229
; Sequence 229, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510.386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 229
; LENGTH: 2602
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (501)..(2102)
US-10-510-386-229
Alignment Scores:
Pred. No.: 9.84e-10 Length: 2602
Score: 157.00 Matches: 32
Percent Similarity: 56.5% Conservatives: 16
Best Local Similarity: 37.6% Mismatches: 33
Query Match: 33.5% Indels: 4
DB: 8 Gaps: 2

US-09-917-376-5 (1-88) x US-10-510-386-229 (1-2602)
Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24
Db 1665 ATGACGATGATACAGAGGGGGACCAATGTAAACGCAACCAATCCGCTCAG 1724
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44

US-09-917-376-5 (1-88) x US-11-179-411-21 (1-1146)
Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24
Db 1 ATGTCAGTTGAATTTTACAACTTAACAAATCAGCAGCAACAACTCAATTCACCAATA 60
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 61 ATCAAAATTAACACACATCTGACAGTCAATTAATTTAAATGAGTAAAGCTTAGATAT 120
```

```
Db 1725 CTCACATTAAAAACACAGCAAAAAACCGTGTCTTTAAATCGAATCACTGTCGCTAC 1784
Qy 45 TrpPhe---ThrArgAspGlyGlySerSerThrLeuValTyrAsnGlyAspTrpAlaAla 63
Db 1785 TGGTATAAACGAATCGCAAGGACAAAT-----TTGACTGCGACTATGCCCAA 1835
Qy 64 MetGlyCysGlyAsnLeuArgAlaSerPheGlySerValAsnProAlaThrProThrAla 83
Db 1836 ATCGGCTGCAGCAAAATCAGCAGCAAAATTCGTTCAATTAATAAAAAAGCGGTAAACGAGCA 1895
Qy 84 AspThrTyrLeuGln 88
Db 1896 GACACGTATCTGAA 1910

RESULT 3
US-11-179-411-21
; Sequence 21, Application US/11179411
; Publication No. US20050266491A1
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: Szczepaniak, William
; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING TH
; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-128
; CURRENT APPLICATION NUMBER: US/11/179,411
; CURRENT FILING DATE: 2005-07-12
; PRIOR APPLICATION NUMBER: US/09/808,898
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 60/189,691
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 08/757,046
; PRIOR FILING DATE: 1996-11-25
; PRIOR APPLICATION NUMBER: 08/597,274
; PRIOR FILING DATE: 1996-02-06
; PRIOR APPLICATION NUMBER: 08/908,909
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: 08/990,103
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Gaussia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1146)
; OTHER INFORMATION: Nucleotide sequence encoding a CBD-Gaussia
; OTHER INFORMATION: luciferase fusion protein
US-11-179-411-21
Alignment Scores:
Pred. No.: 0.00017 Length: 1146
Score: 113.00 Matches: 29
Percent Similarity: 53.2% Conservatives: 21
Best Local Similarity: 30.9% Mismatches: 32
Query Match: 24.1% Indels: 12
DB: 12 Gaps: 4

US-09-917-376-5 (1-88) x US-11-179-411-21 (1-1146)
Qy 5 VallysValGlnTyrLysAsnAspSerAlaProGlyAspAsnGlnLeuLysProGly 24
Db 1 ATGTCAGTTGAATTTTACAACTTAACAAATCAGCAGCAACAACTCAATTCACCAATA 60
Qy 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44
Db 61 ATCAAAATTAACACACATCTGACAGTCAATTAATTTAAATGAGTAAAGCTTAGATAT 120
```



QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaMet 64  
 Db 121 TATTACAAAGTGGTACACAGCAAACTTTCTGG---TGTGACCATGCTGGTGA 177  
 QY 65 GlyCysGlyAsn-----IleArgAlaSerPhe-----Gly 74  
 Db 178 TTATTAGGAATAGTATGTTGATAACACTAGCAAGTACAGCAAACTTCGTTAAAGNA 237  
 QY 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
 Db 238 ACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 276

## RESULT 4

US-11-175-766-21  
 ; Sequence 21, Application US/11175766  
 ; Publication No. US20050272111A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bryan, Bruce  
 ; APPLICANT: Szent-Gyorgyi, Christopher  
 ; APPLICANT: Szczepaniak, William  
 ; TITLE OF INVENTION: RENILLA RENIFORMIS FLOURESCENT PROTEINS, NUCLEIC ACIDS ENCODING T  
 ; TITLE OF INVENTION: FLOURESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIGH TH  
 ; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS  
 ; FILE REFERENCE: 24729-128  
 ; CURRENT APPLICATION NUMBER: US/11/175,766  
 ; CURRENT FILING DATE: 2005-07-06  
 ; PRIOR APPLICATION NUMBER: US/09/808,898  
 ; PRIOR FILING DATE: 2001-03-15  
 ; PRIOR APPLICATION NUMBER: 60/189,691  
 ; PRIOR FILING DATE: 2000-03-15  
 ; PRIOR APPLICATION NUMBER: 09/277,716  
 ; PRIOR FILING DATE: 1999-03-26  
 ; PRIOR APPLICATION NUMBER: 08/757,046  
 ; PRIOR FILING DATE: 1996-11-25  
 ; PRIOR APPLICATION NUMBER: 08/597,274  
 ; PRIOR FILING DATE: 1996-02-06  
 ; PRIOR APPLICATION NUMBER: 08/908,909  
 ; PRIOR FILING DATE: 1997-08-08  
 ; PRIOR APPLICATION NUMBER: 08/990,103  
 ; PRIOR FILING DATE: 1997-12-12  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 21  
 ; LENGTH: 1146  
 ; TYPE: DNA  
 ; ORGANISM: Gaussia  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(1146)  
 ; OTHER INFORMATION: Nucleotide sequence encoding a CBD-Gaussia  
 ; OTHER INFORMATION: luciferase fusion protein  
 US-11-175-766-21

## Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00017 | Length:       | 1146 |
| Score:                 | 113.00  | Matches:      | 29   |
| Percent Similarity:    | 53.2%   | Conservative: | 21   |
| Best Local Similarity: | 30.9%   | Mismatches:   | 32   |
| Query Match:           | 24.1%   | Indels:       | 12   |
| DB:                    | 12      | Gaps:         | 4    |

US-09-917-376-5 (1-88) x US-11-175-766-21 (1-1146)

QY 5 ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsnGlnIleLysProGly 24  
 Db 1 ATGTCAAGTGAATTTTACAACTCTAACAAATCAGCAAAACAACTCAATTACACCAATA 60  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
 Db 61 ATCAAAATTACTAACACATCTGACAGTGAATTTAAATGACGTAAAGTTAGATAT 120  
 QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaMet 64

Db 121 TATTACAAAGTGGTACACAGCAAACTTTCTGG---TGTGACCATGCTGGTGA 177  
 QY 65 GlyCysGlyAsn-----IleArgAlaSerPhe-----Gly 74  
 Db 178 TTATTAGGAATAGTATGTTGATAACACTAGCAAGTACAGCAAACTTCGTTAAAGNA 237  
 QY 75 SerValAsnProAlaThrProThrAlaAspThrTyrLeuGln 88  
 Db 238 ACAGCAAGCCCA---ACATCAACCTATGATACATATGTTGAA 276

## RESULT 5

US-10-517-939-343  
 ; Sequence 343, Application US/10517939  
 ; Publication No. US20060003433A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steer, Brian  
 ; APPLICANT: Callen, Walter  
 ; APPLICANT: Healey, Shaun  
 ; APPLICANT: Hazlewood, Geoff  
 ; APPLICANT: Wu, Di  
 ; APPLICANT: Blum, David  
 ; APPLICANT: Esteghlalian, Alireza  
 ; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM  
 ; TITLE OF INVENTION: AND METHODS FOR MAKING AND USING THEM  
 ; FILE REFERENCE: 564462007901  
 ; CURRENT APPLICATION NUMBER: US/10/517,939  
 ; CURRENT FILING DATE: 2004-12-13  
 ; PRIOR APPLICATION NUMBER: PCT/US03/19153  
 ; PRIOR FILING DATE: 2003-06-16  
 ; PRIOR APPLICATION NUMBER: 60/389,299  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 380  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 343  
 ; LENGTH: 2217  
 ; TYPE: DNA  
 ; ORGANISM: Unknown  
 ; FEATURE:  
 ; OTHER INFORMATION: Obtained from an environmental sample.  
 US-10-517-939-343

## Alignment Scores:

|                        |         |               |      |
|------------------------|---------|---------------|------|
| Pred. No.:             | 0.00159 | Length:       | 2217 |
| Score:                 | 108.00  | Matches:      | 27   |
| Percent Similarity:    | 53.6%   | Conservative: | 18   |
| Best Local Similarity: | 32.1%   | Mismatches:   | 31   |
| Query Match:           | 23.0%   | Indels:       | 8    |
| DB:                    | 8       | Gaps:         | 4    |

US-09-917-376-5 (1-88) x US-10-517-939-343 (1-2217)

QY 7 ValGlnTyrLysAsnAsnAspSerAlaProGlyAspAsn-----GlnIleLysProGly 24  
 Db 1771 CTTCACTACCGC-----TCCGCTGATGGAGATAACAACCTATCAAAATGAAGCCTCAG 1821  
 QY 25 LeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrValArgTyr 44  
 Db 1822 TTCACGATCAAGAACGCGAGCAAGTCCCATCCCGTTAAGCGAGTGACGATCCGCTAC 1881  
 QY 45 TrpPheThrArgAspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaMet 64  
 Db 1882 TATTTACCCCGGAG---AGCAGCGAGCGGTGGATACACGAGTCCGATCGCGGCCAATTC 1938  
 QY 65 GlyCysGlyAsnIleArgAlaSerValAsnProAlaThrProThrAlaMet 84  
 Db 1939 GGAGCAGAGCATGTCACAGACGAGTC-----GTTCCGCCATCCGATCGCGGCGCAC 1992  
 QY 85 ThrTyrLeuGln 88  
 Db 1993 GCCTATGTCGAG 2004

## RESULT 6

[illegible]



```

; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 487
; LENGTH: 4952
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-487

Alignment Scores:
Pred. No.: 470 Length: 4952
Score: 68.50 Matches: 22
Percent Similarity: 44.6% Conservative: 7
Best Local Similarity: 33.8% Mismatches: 25
Query Match: 14.6% Indels: 11
DB: 12 Gaps: 3

US-09-917-376-5 (1-88) x US-11-136-527-487 (1-4952)
QY 31 GlySerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArgAspGly 50
Db 1413 GGGACCCAGAAACAGACCTAGAGTCTAGCTGACGACCTATACACTCAAGGCAGATGGA 1472
QY 51 -----GlySerSer-----ThrLeuValTyrAsnCysAsp----- 60
Db 1473 ACCCAGTGTCCAGTGGTCTTCTCGAAACAACAGCATCTACAGTGTGAGTTCGTGACG 1532
QY 61 ---TrpAlaAlaMetGlyCysGlyAsnIleArgAlaSerPheGlySerValAsnProAla 79
Db 1533 GTCIAGGAGGCAAGGAGGAGTGAAGAACATAGCCGTGACCTTCACCTCTGTAGCCAACTA 1592
QY 80 ThrProThrAlaAsp 84
Db 1593 ACAATAACCCCGGAC 1607

RESULT 12
US-10-517-939-329
; Sequence 329, Application US/10517939
; Publication No. US20060003433A1
; GENERAL INFORMATION:
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Hazlewood, Geoff
; APPLICANT: Wu, Di
; APPLICANT: Blum, David
; APPLICANT: Esteghlalian, Alireza
; TITLE OF INVENTION: XYLANASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 564462007901
; CURRENT APPLICATION NUMBER: US/10/517,939
; PRIOR FILING DATE: 2004-12-13
; PRIOR APPLICATION NUMBER: PCT/US03/19153
; PRIOR FILING DATE: 2003-06-16
; PRIOR APPLICATION NUMBER: 60/389,299
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 329
; LENGTH: 2268
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample.
US-10-517-939-329

Alignment Scores:
Pred. No.: 222 Length: 2268
Score: 68.00 Matches: 31
Percent Similarity: 37.9% Conservative: 8
Best Local Similarity: 30.1% Mismatches: 40
Query Match: 14.5% Indels: 24
DB: 8 Gaps: 5

US-09-917-376-5 (1-88) x US-10-517-939-329 (1-2268)
QY 2 SerGlyGlyVallysValGlnTyrIlys---AsnAsnAspSerAlaProGlyAspAsnGln 20
Db 1951 TCAGGGCGCTGCACGCTGCCTACGCCATCACCACGATCGGGCAGCGT----- 2001
QY 21 IleLysProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrVal 40
Db 2002 TTCACCGCCCAACGTTTACCTTCCCGCAATCAACCATCAGCAACGCCCTGGACGGAACGCC 2055
QY 41 ThrValArgTyrTrpPheThrArgAsp----- 49
Db 2056 ACCCTGGCGCTATGCTTTCCTCCCGCAATCAACCATCAGCAACGCCCTGGACGGAACGCC 2115
QY 50 -----GlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMetGlyCysGly 67
Db 2116 GTTCAGTCCGGCAGCAGCGTTCAGCTCACCACGCCGCTTGAATGGCAGCCTGCCGCC 2175
QY 68 AsnIleArgAlaSerPheGly-----SerValAsnProAlaThrProThr 82
Db 2176 AACGTCTCCGCGAGCTTTGGCTTCCAGGCGAGCTACAGCGGCAATAACAGCGTCCCTGCC 2235
QY 83 AlaAspThr 85
Db 2236 AGCTTTACG 2244

RESULT 13
US-10-995-561-13307/c
; Sequence 13307, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13307
; LENGTH: 35344
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13307

Alignment Scores:
Pred. No.: 5,23e+03 Length: 35344
Score: 68.00 Matches: 16
Percent Similarity: 57.1% Conservative: 4
Best Local Similarity: 45.7% Mismatches: 11
Query Match: 14.5% Indels: 4
DB: 8 Gaps: 1

US-09-917-376-5 (1-88) x US-10-995-561-13307 (1-35344)
QY 51 GlySerSerThrLeuValTyr-----AsnCysAspTrpAlaAlaMetGlyCys 66
Db 23968 GGTAAAGACCACTTCCCTATTTCACCTTGAGGTCAACTGTGACTGGCGCGCTTTTGGC 23909
QY 67 GlyAsnIleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 23908 TTTAGAGTAGGGCATCATTTACCACCTTCTTTCCCTCTAGTCTT 23864

RESULT 14
US-09-925-065A-228751/c
; Sequence 228751, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

```

```
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 228751
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-228751

Alignment Scores:
Pred. No.: 51      Length: 556
Score: 67.50      Matches: 25
Percent Similarity: 37.6%      Conservative: 7
Best Local Similarity: 29.4%      Mismatches: 20
Query Match: 14.4%      Indels: 33
DB: 6      Gaps: 4

US-09-917-376-5 (1-88) x US-09-925-065A-228751 (1-556)

QY 23 ProGlyLeuGlnLeuValAsnThrGlySer----- 32
Db 238 CCTGGACTCCACCTTTGTCATACAGCAGTTAAGGATGCCCTTCTGACAAAGACTTTT 179

QY 33 -----SerSerValAspLeuSerThrValThrValArgTyrTrpPheThrArg 48
Db 178 ATTTTAGCTTAGTTCTCTGTAATTTGTTTGTGAGATTGAGTTACTGTGTAATATGG 119

QY 49 AspGlyGlySerSerThrLeuValTyrAsnCysAspTrpAlaAlaMet----- 64
Db 118 GACCCGGGG-----TAGGACTGGCTGGCGGCTGTCATCCTTCC 80

QY 65 -----GlyCysGlyAsnIleArg---AlaSerPheGlySerVal 76
Db 79 TCCGAGGGCCCTGGGAGAGAGGGTGTGGAGTGGCCGGTGGCTGTTCTCCCCCAATT 20

QY 77 AsnProAlaThrPro 81
Db 19 TCCTGTTGGACCCCT 5

RESULT 15
US-11-072-279/c
; Sequence 279, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 279
; LENGTH: 2564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-279

Alignment Scores:
Pred. No.: 296      Length: 2564
Score: 67.50      Matches: 23
Percent Similarity: 42.1%      Conservative: 9
Best Local Similarity: 30.3%      Mismatches: 27
Query Match: 14.4%      Indels: 17
DB: 3      Gaps: 3

US-09-917-376-5 (1-88) x US-11-072-512-279 (1-2564)

QY 23 ProGlyLeuGlnLeuValAsnThrGlySerSerValAspLeuSerThrValThrVal 42
Db 454 CCTCCGAAGCGATGGCTGACACTCGCTCCAGCAGCTTGTCCGCGAGCTCATCGAATACC 395

QY 43 ArgTyrTrpPheThr----- 47
Db 394 GACTGGTGAATCCAGCCGCGGCTCCCGTCGAGGTCCCAAGAGGGGGGCTGACTGA 335

QY 48 ArgAspGlyGlySerSerThrLeu---ValTyrAsnCysAspTrpAlaAlaMetGlyCys 66
Db 334 AGAGATGGCAGCAGCACACCATTTCTGTCTGGAACCTGCTCGATGGCTTGAGCGGCTCC 275

QY 67 GlyAsn---IleArgAlaSerPheGlySerValAsnProAlaThrPro 81
Db 274 GTGCAGTTGGTCAGGGTCTCTTTCAGGTCTCTCGCCCGTGGCCACGCC 227

Search completed: March 2, 2006, 13:23:29
Job time : 101.433 secs
```